



Determinación del proteoma del dinoflagelado *Pyrocystis lunula* y estudio de sus aplicaciones biotecnológicas

***Memoria presentada por D. Carlos Fajardo Quiñones
para optar al Grado de Doctor en Recursos Agroalimentarios***

Fdo. Carlos Fajardo Quiñones

Los Directores

*Dr. D. Francisco Fernández Acero
Profesor Titular
Universidad de Cádiz*

*Dr. D. Juan Miguel Mancera
Catedrático
Universidad de Cádiz*



El Dr. D. Francisco Javier Fernández Acero, Profesor Titular del área de Microbiología, adscrita al Departamento de Biomedicina, Biotecnología y Salud Pública y el Dr. D. Juan Miguel Mancera Romero, Catedrático de Zoología del Departamento de Biología; ambos de la Facultad de Ciencias del Mar y Ambientales de la Universidad de Cádiz.

INFORMAN: Que la presente memoria titulada *Determinación del proteoma del dinoflagelado Pyrocystis lunula y estudio de sus aplicaciones biotecnológicas* que presenta D. Carlos Fajardo Quiñones para optar al Grado de Doctor en Recursos Agroalimentarios, ha sido realizada bajo su dirección, y autorizan su presentación y defensa.

Y para que así conste y surta los efectos oportunos, firmamos el presente en Puerto Real, a 23 de octubre de 2019.

Dr. D. Francisco Fernández Acero
Profesor Titular
Universidad de Cádiz

Dr. D. Juan Miguel Mancera
Catedrático
Universidad de Cádiz

Esta Tesis Doctoral ha sido realizada gracias a dos becas pre-doctorales, una financiada por la Asociación Universitaria Iberoamericana de Postgrado (AUIP) y la otra por la Oficina de Relaciones Internacionales de la Universidad de Cádiz (ORI-UCA), con una duración global comprendida desde el 04 de octubre de 2015 al 07 de octubre de 2019 (4 años en total), incluyéndose durante dicho periodo cuatro Estancias de Investigación Internacionales (14 meses en total), en el Centro Interdisciplinario de Investigación Marina y Ambiental de la Universidad de Porto (Portugal); Universidad de Ferrara (Italia); Pontificia Universidad Católica de Valparaíso (Chile); y en el Instituto Tecnológico y de Estudios Superiores de Monterrey (México).

Esta Tesis Doctoral se ha desarrollado dentro del marco Campus de Excelencia Internacional en Agroalimentación (CEIA3), habiendo sido parcialmente subvencionada por los Proyectos de Investigación cuyas referencias se detallan a continuación:

- “Producción de Bio-compuestos Saludables en Microalgas (A4HW)”. Ministerio de Economía, Industria y Competitividad, Programa de Colaboración Retos. Fondo Europeo de Desarrollo Regional y ENDESA Generación S.A. (RTC-2016-4860-2).
- “INNOVALGA”. Corporación Tecnológica de Andalucía (Exp. 14/775).
- “Bioluminiscencia & Gastronomía”. Universidad de Cádiz y Mar Cristal Marilum S.L. (OT2016/058).
- “Aplicación de Microorganismos Bioluminiscentes en Alta Cocina. Cultivo y Estudio de su Viabilidad”. Universidad de Cádiz y Aponiente (OT2014/015).

Agradecimientos

Dedicado a mi familia y a todas aquellas personas que me han acompañado y guiado en este viaje existencial, a mis directores/tutores, Dr. Juan Miguel Mancera Romero y Dr. Francisco Javier Fernández Acero. También al Dr. Marcos De Donato, Dr. Gonzalo Martínez Rodríguez y a la Dra. Neda Gilannejad, por ser verdaderos maestros, no solo desde el punto de vista profesional y académico sino también de vida. A mis compañeros de laboratorio y a los amigos que me han apoyado en este largo camino, gracias a todos desde el fondo del alma.

RESUMEN

Se muestra aquí el primer análisis molecular de *Pyrocystis lunula* utilizando enfoques tanto transcriptómicos como proteómicos ("ómico"). *P. lunula* es un dinoflagelado marino que históricamente ha recibido atención debido a su capacidad bioluminiscente según los ritmos circadianos y actualmente existen varias aplicaciones biotecnológicas derivadas de este sistema en esta especie. Por este motivo se empleó un enfoque "ómico" utilizando muestras de *P. lunula* frescas. Como resultado se generaron un total de 80.874.825 lecturas en bruto (11.292.087.505 pb; 55,82 % GC) mediante secuenciación de ARNm. Las secuencias fueron ensambladas en 414.295 contigs (219.203.407 pb; 55,38 % GC) por el software Trinity, generando un transcriptoma de referencia que se empleó posteriormente para llevar a cabo el primer análisis proteómico en *P. lunula*. Se identificaron un total de 17.461 péptidos y 3.182 proteínas. Las proteínas identificadas se categorizaron de acuerdo con la descripción funcional y la clasificación de ontología de genes. Los resultados proteómicos representan una pieza valiosa en la comprensión del mecanismo bioluminiscente de *P. lunula* a nivel molecular y arrojan luz sobre la identificación de factores importantes involucrados en la regulación de la expresión génica. Destaca la presencia de la proteína de unión a luciferina (LBP), que no se había descrito hasta el momento en el género *Pyrocystis*. Los mecanismos celulares y moleculares que subyacen al fenómeno bioluminiscente en dinoflagelados están caracterizados, sin embargo, todavía existen algunos aspectos que siguen siendo un enigma. Tal es el caso de la presencia y diversidad de la proteína "luciferin binding protein" (LBP), así como el proceso de síntesis de luciferina. Estos son factores que son estudiados, estableciéndose nuevas perspectivas en relación con los elementos moleculares responsables de este fenómeno, así como sobre la función de la bioluminiscencia en los dinoflagelados.

ABSTRACT

The first molecular analysis of *Pyrocystis lunula* using both transcriptomic and proteomic approaches ("omic") is reported here. *P. lunula* is a marine dinoflagellate that has historically received attention due to its bioluminescent capacity according to circadian rhythms and there are currently several biotechnological applications derived from this system. For this reason, an "omic" approach was used using fresh *P. lunula* samples. A total of 80,874,825 raw readings (11,292,087,505 bp, 55.82 % GC) were generated by mRNA sequencing. The sequences were assembled into 414,295 contigs (219,203,407 bp, 55.38 % GC) by the Trinity software, generating a reference transcriptome that was subsequently used to carry out the first proteomic analysis in *P. lunula*. A total of 17,461 peptides and 3,182 proteins were identified. The identified proteins were categorized according to the functional description and gene ontology classification. The proteomic results represent a valuable piece in the understanding of the bioluminescent mechanism of *P. lunula* at the molecular level and sheds light on the identification of important factors involved in the regulation of gene expression. It highlights the presence of the luciferin binding protein (LBP), which had not been described so far in the genus *Pyrocystis*. The cellular and molecular mechanisms that underlie the bioluminescent phenomenon in dinoflagellates are well characterized; however, there are still some aspects that remain as an enigma. This is the case of the presence and diversity of LBP, as well as the synthesis process of luciferin. These factors are systematically studied, setting new perspectives in relation to the molecular elements responsible for this phenomenon, as well as the function of bioluminescence in dinoflagellates.

ÍNDICE GENERAL

	Página
1. INTRODUCCIÓN	11
1.1 EL GÉNERO PYROCYSTIS	12
1.2 BIOLUMINISCENCIA EN DINOFLAGELADOS	13
1.2.1 Bases bioquímicas de la bioluminiscencia	14
1.2.2 Luciferasa (LCF)	17
1.2.3 Proteína de unión a la luciferina (LBP)	19
1.2.4 Luciferina	20
1.3 FUNCIÓN DE LA BIOLUMINISCENCIA EN DINOFLAGELADOS	22
1.4 RETOS EN EL ESTUDIO PROTEÓMICO DE DINOFLAGELADOS	24
1.4.1 Aplicación de tecnologías moleculares de alto rendimiento: transcriptómica y proteómica	26
1.5 APLICACIONES BIOTECNOLÓGICAS DE LAS MICROALGAS	29
1.6 INGENIERÍA GENÉTICA EN MICROALGAS	32
1.7 HIPÓTESIS Y OBJETIVOS	35
2. MATERIALES Y MÉTODOS	36
2.1 CONDICIONES DE CULTIVO	37
2.2 AISLAMIENTO DE ARN TOTAL	37
2.3 SECUENCIACIÓN DE TRANSCRIPTOMA	38
2.4 ENSAMBLAJE <i>DE NOVO</i> Y ANÁLISIS DE SECUENCIACIÓN DE ARN DE ALTO RENDIMIENTO	38
2.5 ANÁLISIS PROTEÓMICO	40
2.6 ONTOLOGÍA GÉNICA Y ANÁLISIS KEGG	41
2.7 VERIFICACIÓN DE LA PRESENCIA DE LBP Y LCF / LBP POR PCR	42
2.8 ANÁLISIS FILOGENÉTICOS Y ESTRUCTURALES	43
3. RESULTADOS	45
3.1 ENSAMBLAJE DE TRANSCRIPTOMA <i>DE NOVO</i>	46

3.2 ANÁLISIS PROTEÓMICO	50
3.3 ANOTACIÓN FUNCIONAL	51
4. DISCUSIÓN	56
4.1 ANÁLISIS ESTRUCTURALES Y FILOGENÉTICOS	57
4.1.1 Luciferasa (LCF)	57
4.2 NUEVAS PROTEÍNAS Y MINERÍA DE DATOS PROTEÓMICOS	64
4.2.1 Proteína de unión a la luciferina (LBP)	65
4.2.2 Glutación-S-transferasa (GST) y la escurridiza luciferina	70
4.3 ESTUDIO DE POSIBLES APLICACIONES BIOTECNOLÓGICAS	77
4.3.1 Luciferasa (LCF)	78
4.3.2 Glutación-S-transferasa (GST)	80
4.3.3 Transferencia de Carga Intramolecular Retorcida (TICT)	82
4.4 PERSPECTIVAS FUTURAS	83
5. CONCLUSIONES	85
6. REFERENCIAS	89
7. MATERIAL SUPLEMENTARIO	101
ANEXOS	103

ÍNDICE DE FIGURAS

	Página
Figura 1. Célula vegetativa de <i>P. lunula</i> (Schutt).	13
Figura 2. Mecanismo de mecano-transducción responsable del evento bioluminiscente en dinoflagelados.	15
Figura 3. Reacción bioluminiscente en dinoflagelados (LH2 = luciferina).	15
Figura 4. Estructura de la molécula de luciferina aislada de <i>P. lunula</i> .	21
Figura 5. Comparación del análisis estructural de 20 ensamblajes transcriptómicos generados por RAY.	48
Figura 6. Comparación del análisis estructural de 10 ensamblajes transcriptómicos generados por los tres algoritmos (RAY, Trinity, MIRA).	48
Figura 7. Análisis de representatividad por similitud de secuencia del conjunto Trinity_Q20 L100_k25_NC20 (líneas rojas), contra los 25 conjuntos, así como dos conjuntos de datos públicos de <i>P. lunula</i> (SRA: SRS621557) y <i>B. minutum</i> (SRA: DRS003180).	49
Figura 8. Distribución de las proteínas predichas e identificadas de acuerdo con su peso molecular (MW) y el punto isoelectrico (pI) calculado por los péptidos de la biblioteca R.	52
Figura 9. Análisis funcional de GO de las secuencias de <i>P. lunula</i> según la clasificación del proceso biológico (BP).	53
Figura 10. Análisis funcional de GO de las secuencias de <i>P. lunula</i> según la clasificación de la función molecular (MF).	54
Figura 11. Distribución de proteínas por especies para las secuencias de <i>P. lunula</i> según BLAST2GO.	55
Figura 12. (A) Secuencia de la proteína LCF A de <i>P. lunula</i> (GenBank AAL40676) que muestra los dominios LCF / LBP N-terminal, paquete helicoidal y catalítico, así como las dos regiones conservadas dentro del dominio catalítico. (B) Estructura 2D / 3D prevista del haz helicoidal (amarillo) y dominio catalítico de la enzima luciferasa (azul) de <i>P. lunula</i> (GenBank AAL40676), que muestra las regiones conservadas 1 (rojo) y 2 (verde), hechas con Servidor web Phyre2 y visualizado con EzMol.	58
Figura 13. La alineación de la secuencia de las regiones conservadas 1 (A) y 2 (B) encontradas en la enzima LCF se compara entre los tres dominios catalíticos y las secuencias disponibles de las especies de <i>Alexandrium</i> , <i>Lingulodinium</i> , <i>Pyrocystis</i> y <i>Protoceratium</i> .	59
Figura 14. Análisis filogenético molecular por un método bayesiano, que muestra el patrón evolutivo de los tres dominios catalíticos de LCF en las especies de dinoflagelados con secuencias de proteínas disponibles, utilizando el único dominio catalítico de <i>N. scintillans</i> como grupo externo.	60

Figura 15. Árbol filogenético de los dinoflagelados Gonyaulacales inferidos de los tres genes de ADNr, dos genes mitocondriales y cuatro proteínas nucleares, modificados de Orr <i>et al.</i> (2012).	61
Figura 16. Alineación de secuencia de los tres dominios del paquete helicoidal contenidos en la enzima LCF de las especies de Alexandrium, Lingulodinium, Ceratocorys, Pyrocystis y Protoceratium que tienen su proteína publicada en GenBank.	62
Figura 17. Alineación de secuencia del dominio terminal LCF / LBP N contenido en la enzima LCF, LBP y GST de las especies de Alexandrium, Lingulodinium, Pyrocystis y Protoceratium que tienen su proteína publicada en GenBank.	69
Figura 18. Análisis filogenético molecular por un método bayesiano, que muestra el patrón evolutivo de la proteína LBP de las especies de Alexandrium, Lingulodinium, Pyrocystis y Protoceratium que tienen su proteína publicada en GenBank, utilizando la secuencia de <i>N. scintillans</i> como grupo externo.	70
Figura 19. Reacción enzimática acoplada responsable de la bioluminiscencia de <i>P. lunula</i> .	74

ÍNDICE DE TABLAS

	Página
Tabla 1. Especies de microalgas con estudios proteómicos hasta la fecha.	28
Tabla 2. Números de acceso de las secuencias reportadas en GenBank para la luciferasa (LCF), la proteína de unión a luciferina (LBP) y la glutatión S-transferasa (GST) en microalgas.	44
Tabla 3. Absorción, excitación de fluorescencia y emisión máxima (nm) de P630 oxidado, P630 reducido y luciferina en 10 mM Hepes-NaOH buffer a pH 7.2.	74

1. INTRODUCCIÓN

Los dinoflagelados son un grupo de eucariotas unicelulares que viven en los ecosistemas acuáticos, marinos y de agua dulce, cuyo tamaño varía entre 10 y 100 μm . Son los productores primarios eucarióticos más importantes del océano, solo superados por las diatomeas (Muscatine *et al.*, 1981). Los dinoflagelados son los responsables de las floraciones de algas llamadas comúnmente “mareas rojas”; y más de sesenta especies de estos organismos producen toxinas que tienen un profundo impacto en la industria pesquera, las zonas costeras y la salud pública. Sin embargo, también desempeñan un papel clave en el mantenimiento de la biodiversidad que rodea los arrecifes de coral, ya que los pólipos de coral dependen de los productos fotosintéticos suministrados por los dinoflagelados simbióticos que albergan para crecer en aguas pobres en nutrientes (Lin, 2011; Carrasco *et al.*, 2018; Fajardo *et al.*, 2019). Además de esto, la bioluminiscencia de muchos dinoflagelados ha inspirado no solo el arte y la literatura, sino también la intensiva disección científica de este fenómeno (Schmitter *et al.*, 1976; Herring, 1987; Li y Hastings, 1998; Shimomura, 2006; Haddock *et al.*, 2010; Craig y Pride, 2012; Valiadi e Iglesias, 2013).

1.1 EL GÉNERO PYROCYSTIS

Desde un punto de vista taxonómico y según el NCBI, la clasificación de *Pyrocystis lunula* (Schutt) es la siguiente:

Dominio: Eukaryota
Superfilo: Alveolata
Clase: Dinophyceae
Orden: Pyrocystales
Familia: Pyrocystaceae

Pyrocystis es un género foto-autotrófico que ha sido estudiado desde un punto de vista bioquímico como organismo modelo debido a su capacidad de producir bioluminiscencia según los ritmos circadianos (Seo y Fritz, 2000; Heimann *et al.*, 2009). El género *Pyrocystis* incluye un pequeño número de especies marinas que presentan una distribución global en todos los océanos del mundo. Este género es, desde un punto de vista morfológico, inusual entre los Dinophyceae pues pasa gran parte de su ciclo de vida como una célula no

móvil dentro de una cubierta de celulosa (Swift y Remsen, 1970; Steidinger y Tangen, 1997; Seo y Fritz, 2001). Una medición de los contenidos de ADN en diferentes etapas del ciclo celular de *P. lunula* ha demostrado que la síntesis de ADN es discreta, similar a la de otros dinoflagelados móviles. Las células vegetativas de *P. lunula* (Fig. 1) son diploides (2n) con 49 cromosomas (98 en total) (Seo y Fritz, 2006).



Figura 1. Célula vegetativa de *P. lunula* (Schutt), longitud 100-140 μm (Fuente: AlgaeBase).

1.2 BIOLUMINISCENCIA EN DINOFLAGELADOS

La bioluminiscencia hace referencia a la capacidad de los organismos vivos de producir luz. Desde un punto de vista etimológico, el término se origina del griego *βίος* que significa "vida" y del latín *lumen*, "luz". Este fenómeno no es común entre los organismos terrestres y está prácticamente ausente en organismos de ambientes de agua dulce. De este modo, la mayoría de los organismos bioluminiscentes viven en el medio marino, desde el ecuador hasta las regiones polares y desde la costa hasta las profundidades del océano. Una amplia gama de organismos marinos es capaz de producir este fenómeno, incluyendo bacterias, protistas unicelulares, celentéreos, anélidos, moluscos, crustáceos, equinodermos, peces y medusas, entre otros (Herring, 1987; Haddock *et al.*, 2010).

1.2.1 Bases bioquímicas de la bioluminiscencia

La bioluminiscencia se produce por una reacción enzimática, y aunque el proceso no es el mismo para todos los organismos bioluminiscentes, todos a su vez comparten la misma reacción base. En esta reacción base, la luciferina (sustrato) reacciona con oxígeno en presencia de la enzima luciferasa (LCF) y como resultado se produce una oxiluciferina que emite fotones a medida que decae desde un estado de alta energía a un estado de baja energía. La bioluminiscencia puede ser provocada por estimulación mecánica, eléctrica o química. El tipo y la fuerza de la estimulación afecta a la intensidad y la duración del evento bioluminiscente resultante (Craig y Priede, 2012).

Los dinoflagelados son los principales protistas eucariotas capaces de producir luz (Widder, 2001; Haddock *et al.*, 2010), siendo la bioluminiscencia descrita en aproximadamente el 6 % de los géneros, todos ellos marinos (Hastings y Morin, 1991). El sistema de bioluminiscencia de los dinoflagelados es similar entre especies con respecto a las reacciones cruzadas entre enzimas y sustratos, perfiles de actividad, de pH y su localización celular en orgánulos emisores de luz llamados escintilos (Schmitter *et al.*, 1976; Nicolas *et al.*, 1987). Al igual que en todos los sistemas de bioluminiscencia, el de los dinoflagelados es único desde una perspectiva tanto celular como molecular. Los escintilos (DeSa y Hastings, 1968) contienen el sustrato luciferina, la enzima LCF y, en algunas especies, una proteína de unión a la luciferina (LBP). (Schmitter *et al.*, 1976; Johnson *et al.*, 1985; Knaust *et al.*, 1998). Los escintilos son vesículas densas de aproximadamente 0.5 a 0.9 μm de diámetro (Johnson *et al.*, 1985; Nicolas *et al.*, 1987) que, durante las horas de oscuridad, abundan en la periferia de la célula (Seo y Fritz, 2000). Un destello de luz es activado por un potencial de acción a lo largo de la membrana de la vacuola, inducido por la agitación mecánica de la célula, con la participación de un canal de protones dependiente de voltaje (Smith *et al.*, 2011). Esto conduce a una reducción en el pH dentro de los escintilos, lo que activa a la LCF y provoca que la LBP libere la luciferina, haciéndola disponible para la oxidación por la LCF (Fogel y Hastings 1971) (Fig. 2 y 3). *P. lunula* emite más luz que la mayoría de los otros dinoflagelados bioluminiscentes. Tras la estimulación completa durante la

noche, una única célula de *P. lunula* emite aproximadamente 40×10^8 fotones, lo cuales un orden de magnitud mayor que la luz emitida por *Pyrodinium bahamense* y *Lingulodinium polyedrum* (Valiadi e Iglesias, 2013).

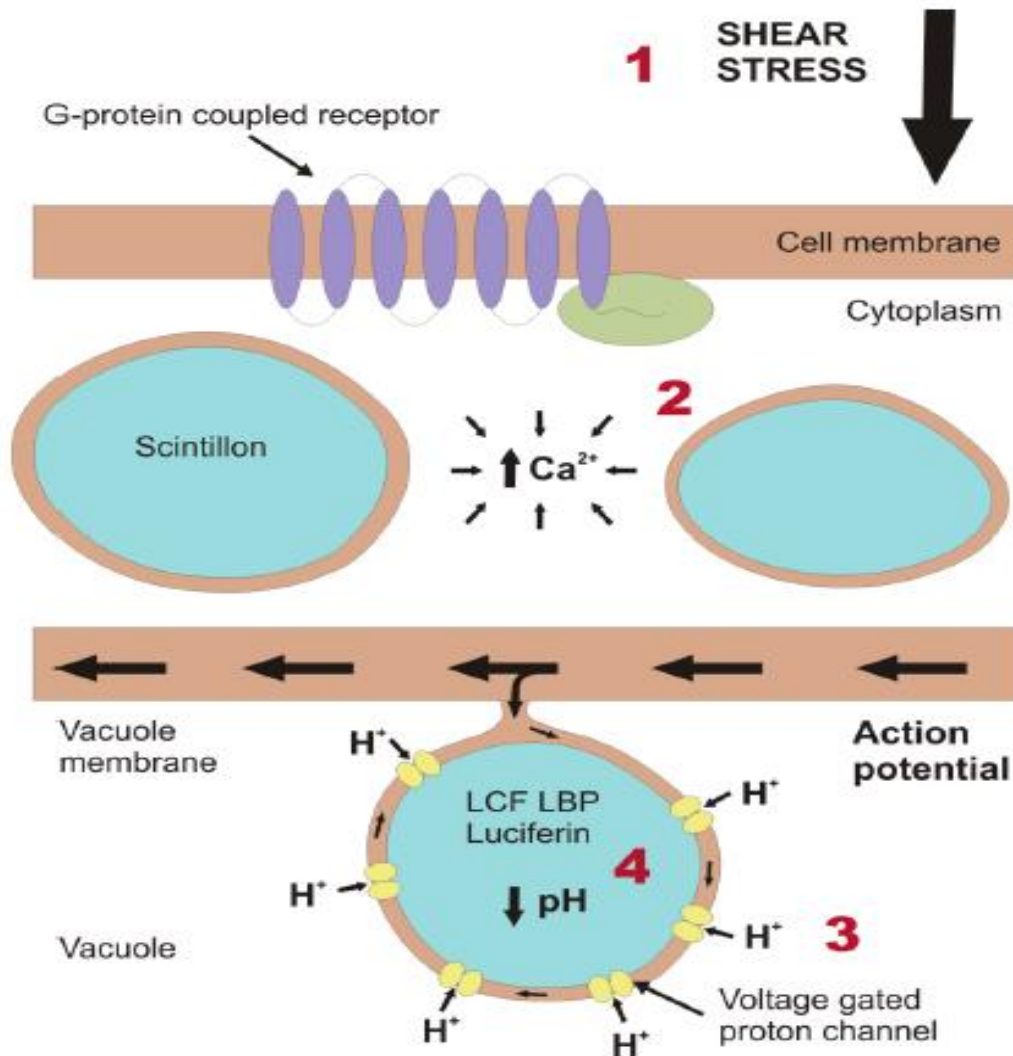


Figura 2. Mecanismo de mecano-transducción responsable del evento bioluminiscente en dinoflagelados (Fuente: Valiadi e Iglesias, 2013).

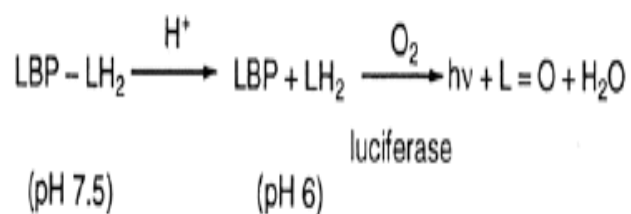


Figura 3. Reacción bioluminiscente en dinoflagelados (LH₂ = luciferina) (Fuente: Haddock *et al.*, 2010).

La bioluminiscencia representa uno de los sistemas de mecanosensibilidad más rápidos conocidos hasta la fecha, con un retraso de solo 15 a 20 ms entre el estímulo y la respuesta (Eckert, 1965; Widder y Case, 1981; Latz *et al.*, 2008). En *P. lunula*, la rigidez celular es de 0.56 MPa, consistente con células que poseen una pared celular. La respuesta celular depende tanto de la magnitud como de la velocidad de la fuerza aplicada. A la velocidad de estimulación máxima de $390 \mu\text{m s}^{-1}$, la respuesta del umbral se produce a una fuerza de 7.2 μN , lo que resulta en un tiempo de contacto de 6.1 ms y una indentación de 2.1 μm . Las células no responden a una velocidad de estimulación baja de $20 \mu\text{m s}^{-1}$, lo que indica una respuesta dependiente de la velocidad que, según los experimentos de relajación de estrés, se explica por las propiedades viscoelásticas de las células (Tesson y Latz, 2015).

Un fenómeno mucho menos investigado que también muestra control circadiano es la bioluminiscencia espontánea. Esto se ha registrado en *L. polyedrum* (Krasnow *et al.*, 1980), *P. lunula* (Biggley *et al.*, 1969; Colepicolo *et al.*, 1993), *Pyrodinium bahamense* (Biggley *et al.*, 1969) y *Ceratocorys horrida* (Latz y Lee, 1995). *L. polyedrum* emite destellos espontáneos al comienzo de la fase oscura y un brillo continuo al final de la fase oscura. *P. lunula* y *P. bahamense* emiten destellos espontáneos durante toda la fase oscura, pero no generan un brillo significativo. Se ha observado una emisión de luz constante de bajo nivel en las tres especies durante la fase oscura. *C. horrida* emite destellos espontáneos a lo largo de la fase oscura y un brillo al final. Es probable que las emisiones espontáneas de destello y brillo y el bajo nivel de fondo representen procesos bioquímicos y celulares que podrían arrojar luz sobre la función endógena (es decir, no estimulada externamente) de la bioluminiscencia (Valiadi e Iglesias, 2013). También se ha observado una variación intraespecífica en la presencia/ausencia e intensidad de la bioluminiscencia en cepas de una sola población natural de *P. lunula* (Swift *et al.*, 1973). Por lo tanto, las características de la bioluminiscencia producida por cada especie, o incluso por cada célula, pueden diferir significativamente debido a varios factores de control endógenos (Valiadi e Iglesias, 2013).

Los mecanismos celulares de la bioluminiscencia en dinoflagelados y la evolución molecular de los genes involucrados en el proceso están bien caracterizados (genes LCF y de la proteína de unión a la luciferina o LBP). Los aspectos más destacados de los últimos 25 años han sido la secuenciación de los genes que codifican a la LCF de varias especies, especialmente *Noctiluca scintillans*, así como la caracterización de la luciferina de *P. lunula*. Estos hitos han constituido avances significativos en nuestra comprensión de los componentes celulares y vías de mecano-transducción que inducen la producción de luz. La recopilación de datos sobre estas especies y los primeros experimentos ecológicos que implican dinoflagelados bioluminiscentes también han dado lugar a información valiosa sobre el tema (Valiadi e Iglesias, 2013).

1.2.2 Luciferasa (LCF)

De entre los genes mejor caracterizados en dinoflagelados, son posiblemente aquellos relativos al sistema de bioluminiscencia. En *P. lunula*, la LCF cataliza la oxidación de la luciferina de dinoflagelado por el oxígeno molecular, dando como resultado una especie excitada electrónicamente que emite luz azul ($\lambda_{\text{max}} = 480 \text{ nm}$). Esta LCF tiene tres dominios catalíticos en su única cadena polipeptídica ($M_r = \text{ca. } 137 \text{ kDa}$), cada dominio (aproximadamente 46 kDa) es enzimáticamente activo (Morishita *et al.*, 2002; Liu *et al.*, 2004) y su gen de codificación se produce en varias copias agrupadas en tándem dentro del genoma. Los altos números de copias de LCF pueden haber resultado del requerimiento de una gran abundancia de esta proteína destinada a la bioluminiscencia, lo que ejerce una presión selectiva para la retención de copias genéticas duplicadas (Li *et al.*, 1997; Lin, 2011).

La caracterización de las 3 isoformas del gen LCF en la especie *P. lunula* LCFa (GenBank AF394059.1), LCFb (GenBank AF394060.1), y LCFc (GenBank AF394061.1) (Okamoto *et al.*, 2001) revelan que la estructura primaria de LCF en dinoflagelados está muy conservada. La similitud de secuencia y los análisis filogenéticos de los dominios de LCF de *P. lunula* y *L. polyedrum* proporcionan evidencia del origen de esta estructura genética única antes de la divergencia de estos géneros durante la evolución. Más llamativo

es el hallazgo de que, si bien la organización de tres dominios repetidos intramolecularmente está presente en los genes LCF de ambos dinoflagelados, existen diferencias sustanciales en la frecuencia de sustituciones sinónimas a lo largo de sus dominios repetidos. La mayor conservación en el centro de los dominios de la LCF observada en ambas especies sugiere que esta región contiene los sitios activos catalíticos. Esto se confirma por el hecho de que, en *P. lunula*, los dominios de LCF individuales, así como las moléculas de longitud completa, son enzimáticamente activos cuando se expresan en *E. coli*. En *L. polyedrum*, se informó que las actividades correspondientes eran máximas a pH 6.3, disminuyendo casi a cero a pH 8, y cuatro histidinas conservadas intramolecularmente están involucradas en la pérdida de actividad a pH alto. Estos residuos de histidina también se conservan en *P. lunula*. La frecuencia de las sustituciones sinónimas en *P. lunula* es uniforme a lo largo de los dominios (Li *et al.*, 1997; Okamoto *et al.*, 2001; Morishita *et al.*, 2002; Liu *et al.*, 2004).

En algunos casos, la abundancia de proteína LCF exhibe un ritmo circadiano, se sintetiza rápidamente en la oscuridad temprana y se destruye en el período de luz. Sin embargo, existen claras diferencias en el sistema de *P. lunula*, donde el parpadeo, pero no el brillo tiene un ritmo circadiano. Además, las cantidades de luciferina y de LCF son constantes durante el día y la noche (Colepicolo *et al.*, 1993; Knaust *et al.*, 1998); por lo tanto, en lugar de la síntesis diaria *de novo* y la destrucción de todos los componentes, el ritmo parece implicar cambios en su localización intracelular. En este caso, los escintilos son relocalizados de forma intercambiable con los cloroplastos, de forma tal que durante la noche los escintilos se ubican en la periferia de la célula y migran hacia el centro de la misma durante el día. De esta forma pueden modular la intensidad bioluminiscente durante el día al impedir su estimulación (Widder y Case, 1982; Seo y Fritz, 2000). Aunque no se ha determinado el tamaño del genoma de *P. lunula*, se sabe que los dinoflagelados tienen genomas grandes que varían de 2 a 200 pg / núcleo (Spector, 1984). Dentro de estos límites, 200 es el número mínimo de copias de secuencias de LCF en *P. lunula*, el doble del número estimado para *L. polyedrum* (Li y Hastings, 1998). Esta estimación parece razonable, ya que se ha reportado que la cantidad del sustrato luciferina es 100 veces mayor en *P. lunula* que en *L. polyedrum* (Schmitter *et al.*, 1976).

1.2.3 Proteína de unión a la luciferina (LBP)

La emisión de luz en dinoflagelados depende de la oxidación catalizada por una LCF de un sustrato (luciferina) de alta energía, normalmente secuestrado y protegido de la oxidación por una LBP. La LBP desempeña el importante papel de unirse a la luciferina a pH fisiológico, pero ha recibido mucha menos atención que la LCF. Una razón para esto podría ser que no se ha encontrado en todos los dinoflagelados bioluminiscentes y, por lo tanto, no se ha considerado como un componente esencial del sistema de bioluminiscencia de los dinoflagelados. La estructura física de la LBP aún no se ha caracterizado, por lo que no está claro cómo se une exactamente a la luciferina, pero se cree que existe como un dímero que se une a una molécula de luciferina y la libera por cambio conformacional por debajo de pH 7 (Valiadi e Iglesias, 2013; 2014). La especie de dinoflagelado *L. polyedrum* (anteriormente *Gonyaulax polyedra*) ha sido el principal organismo modelo para los estudios de LBP, y representa el primer organismo del cual esta proteína fue aislada, clonada y secuenciada (Morse *et al.*, 1989; Lee *et al.*, 1993; Machabee *et al.*, 1994). La familia de genes LBP de *L. polyedrum* está compuesta por dos tipos de genes, LBP_a y LBP_b, que comparten una identidad de secuencia del 86 % y codifican dos isoformas de proteínas que se expresan en cantidades iguales (Morse *et al.*, 1989; Machabee *et al.*, 1994). Además, cada tipo de gen sigue la organización genómica de dinoflagelado típica y está presente en varias copias no idénticas dispuestas en tándem dentro del genoma (Machabee *et al.*, 1994; Tanikawa *et al.*, 2004).

La presencia del gen LBP es bien conocida en *L. polyedrum* (Fogel *et al.*, 1971; Schmitter *et al.*, 1976), *N. scintillans* (Liu y Hastings, 2007). También se ha encontrado en los transcriptomas de cuatro especies de *Alexandrium* (Hackett *et al.*, 2005; Erdner y Anderson, 2006; Uribe *et al.*, 2008; Jaeckisch *et al.*, 2011), y en los géneros *Ceratocorys*, *Gonyaulax* y *Protoceratium*. Asimismo, se ha demostrado que la LBP es muy abundante (1 %) en el proteoma de *L. polyedrum* (Valiadi e Iglesias, 2013; 2014). Estudios transcriptómicos de *Alexandrium catenella* (Uribe *et al.*, 2008; Toulza *et al.*, 2010) y *L. polyedrum* (Tanikawa *et al.*, 2004) sugieren que la LBP se produce a una tasa mayor que

la LCF. Del mismo modo, la concentración de LCF se estima que es aproximadamente 100 veces menor que el valor informado para LBP (Dunlap y Hastings, 1981). Esto podría explicarse si la LCF se puede reciclar en la reacción bioluminiscente, mientras que la LBP es proporcional a la cantidad de luciferina presente (Valiadi e Iglesias, 2013).

Estos datos indican claramente que, en organismos donde está presente, la LBP es una parte central del sistema de bioluminiscencia y, por lo tanto, se necesitan más estudios sobre la identidad y el origen de esta proteína ya que parece ser un componente integral de la maquinaria molecular de bioluminiscencia estándar en los dinoflagelados. Las secuencias de nucleótidos de LBP mostraron una alta variación intra-cepas, revelando una familia de genes muy diversa que comprende múltiples tipos de genes en algunos organismos (Valiadi e Iglesias, 2013; 2014). De hecho, uno de los únicos géneros de dinoflagelados bioluminiscentes del que no se ha encontrado la expresión de LBP es *Pyrocystis*, cuyos extractos de proteínas se han analizado con un anticuerpo universal para LBP (Schmitter *et al.*, 1976; Knaust *et al.*, 1998) y también a través de pruebas de amplificación mediante PCR (Valiadi e Iglesias, 2014). Actualmente se desconoce la cantidad de especies de dinoflagelados bioluminiscentes que utilizan LBP (Valiadi e Iglesias, 2013). Sin embargo, información emergente muestra evidencia sustancial de que la LBP, con su compleja historia evolutiva, es una parte importante de los sistemas de bioluminiscencia dinoflagelados (Valiadi e Iglesias, 2014).

1.2.4 Luciferina

La estructura de la molécula de luciferina solo se ha caracterizado en *P. lunula* debido en parte a que las células carecen de una teca, y porque contienen aproximadamente 100 veces más luciferina que *L. polyedrum* (Schmitter *et al.*, 1976), más que lo reportado en ningún otro dinoflagelado hasta la fecha. Esta luciferina es extremadamente lábil, ante el pH, la fotooxidación, a una alta concentración de sal y al O₂ (Dunlap y Hastings, 1981). La estructura de la luciferina de *P. lunula* fue completamente caracterizada por Nakamura *et al.* (1989) como un tetrapirrol lineal con similitud

a la clorofila a (Fig. 4). Esta molécula exhibe una fluorescencia azul característica (475 nm) cuando se irradia con luz UV, una característica que también se usa para visualizar los escintilos en las células mediante microscopía de epi-fluorescencia; sin embargo, la fluorescencia de la molécula cesa cuando se ha oxidado por la LCF o por el aire (Valiadi e Iglesias, 2013; 2014). En general, se asume que todas las luciferinas de los dinoflagelados son iguales a las de *P. lunula*, ya que se ha demostrado que su luciferina reacciona de forma cruzada con la LCF de todas las demás especies de dinoflagelados bioluminiscentes (Hamman y Seliger, 1972; Schmitter *et al.*, 1976).

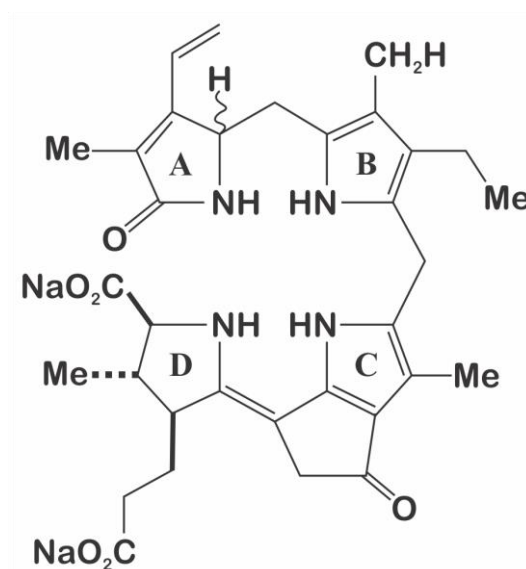


Figura 4. Estructura de la molécula de luciferina aislada de *P. lunula* (Fuente: Nakamura *et al.*, 1989).

Sorprendentemente, este también es el caso del camarón euphasido (krill), *Euphasia superba*, la LCF y luciferina de esta especie también puede reaccionar de forma cruzada con la luciferina y la LCF de *P. lunula*. De hecho, la luciferina de *E. superba* es muy similar a la de *P. lunula*, pero está más estrechamente relacionada con la clorofila c (Topalov *et al.*, 2001), y presumiblemente es obtenida mediante la ingestión de sus presas (Valiadi e Iglesias, 2013). Es importante tener en cuenta que, si bien las luciferinas y las LCFs reaccionan de forma cruzada incluso entre organismos filogenéticamente muy distantes, como los dinoflagelados y el krill, sus LCFs presentan características completamente diferentes. Por ejemplo, la LCF de euphasidos

no se ha clonado, pero su peso molecular es de 600 kDa (Shimomura, 1995), en contraste con los aproximados 135 kDa de las LCFs en dinoflagelados fotosintéticos (Li *et al.*, 1997; Liu *et al.*, 2004).

Existe cierta controversia en relación con el proceso de síntesis de la luciferina, por un lado, un estudio con trazadores de aminoácidos ha confirmado la producción intracelular de luciferina en *P. lunula* (Wu *et al.*, 2003), mientras que Topalov y Kishi (2001) sugirieron que la luciferina es un producto de la degradación de la clorofila por medio de la fotooxidación. Según este último paradigma, Liu y Hastings (2007) plantearon la hipótesis de que los dinoflagelados heterótrofos adquieren luciferina nutricionalmente de las presas ingeridas, ya sea directamente o por la degradación de la clorofila de la presa; sin embargo, el dinoflagelado heterotrófico *Protoperdinium crassipes* puede mantener su intensidad de bioluminiscencia durante 1 año en ausencia de alimentos que contengan clorofila o luciferina (Yamaguchi y Horiguchi 2008) y, por lo tanto, debe contener una luciferina procedente de una molécula precursora endógena diferente (Valiadi e Iglesias, 2013). Además, *L. polyedrum* solo contiene luciferina por la noche cuando no es posible la fotooxidación. La presencia de LBP en varias especies fotosintéticas sugiere que incluso dentro de estas especies, podría haber una molécula de luciferina alternativa que requiera LBP para la estabilización (Morse *et al.*, 1989).

1.3 FUNCIÓN DE LA BIOLUMINISCENCIA EN DINOFLAGELADOS

En el caso de los dinoflagelados, se ha propuesto que la defensa contra los depredadores es el principal papel ecológico de la bioluminiscencia (Abrahams y Townsend, 1993; Buskey *et al.*, 1983; Marcinko *et al.*, 2013) y, por lo tanto, tendría una función fundamental en la estructura del ecosistema. En el medio marino, el papel de la bioluminiscencia se ha estudiado principalmente en bacterias y en la mega fauna de aguas profundas, donde desempeña varias funciones importantes como el camuflaje, el cortejo y la defensa contra los depredadores (Haddock *et al.*, 2010; Widder, 2010); sin embargo, el papel de la bioluminiscencia en los dinoflagelados se ha estudiado de manera menos

detallada y algunos de los conceptos que se han sugerido solo están respaldados por información experimental insuficiente.

Se ha informado que el mecanismo de bioluminiscencia de los dinoflagelados puede ser activado por la proximidad o el contacto con el zooplancton (Latz *et al.*, 1994; Maldonado *et al.*, 2007). Por otro lado, varios estudios han reportado las respuestas fotofóbicas del zooplancton marino al destello de luz artificial (Buskey *et al.*, 1983; Buskey *et al.*, 1987), lo que sugiere que la bioluminiscencia podría otorgar una ventaja evolutiva a los dinoflagelados al reducir la presión selectiva de depredación. Una teoría ampliamente aceptada propuesta hace mucho tiempo es que la bioluminiscencia protege los dinoflagelados indirectamente al actuar como una "alarma antirrobo", llamando la atención de los depredadores visuales de orden superior sobre la ubicación del copépodo (Burkenroad, 1947). Se postuló que incluso si el dinoflagelado intermitente muere, la población continuará beneficiándose porque las floraciones se forman de manera asexual (Valiadi e Iglesias, 2013). Se ha informado que las copepodamidas, lípidos polares exudados por los copépodos, inducen un aumento en la bioluminiscencia de *L. polyedra*, generando un destello más brillante (Prevett *et al.*, 2019). Además, el copépodo *Acartia tonsa* pasa de preferir *L. polyedra* como presa cuando no produce bioluminiscencia, a un rechazo casi completo cuando *L. polyedra* se trata previamente con copepodamidas para inducir una mayor capacidad bioluminiscente (Lindström *et al.*, 2017).

Hanley y Widder (2017) propusieron tres hipótesis para explicar por qué la bioluminiscencia de dinoflagelados disuade la depredación por copépodos: advertencia aposemática, respuesta de sobresalto y alarma antirrobo. La alarma antirrobo es la hipótesis más aceptada; sin embargo, exige una alta concentración en la población bioluminiscente para ser eficaz, por lo tanto, la ventaja selectiva de bioluminiscencia a concentraciones más bajas podría ser el resultado de otro papel como la advertencia aposemática o la respuesta de sobresalto.

Por otro lado, Wilson y Hastings (2013) han propuesto otra hipótesis para el papel original de la bioluminiscencia. Esto se basa en los sistemas de bioluminiscencia de luciérnagas y bacterias; sin embargo, también es factible para dinoflagelados. En esta, llamada "hipótesis de defensa del oxígeno", argumentan que los sistemas de bioluminiscencia evolucionaron en respuesta a bajos niveles de oxígeno durante el gran evento de oxidación, cuando evolucionó la fotosíntesis. Dado que todos los sistemas de bioluminiscencia consumen oxígeno, estos autores señalan que pueden desintoxicar de especies reactivas de oxígeno como peróxidos inorgánicos y orgánicos, radicales libres, iones de oxígeno, y la producción de luz es solo un subproducto. La bioluminiscencia habría obtenido una función diferente cuando las vías antioxidantes se generalizaron al aumentar los niveles de oxígeno.

Según Fresneau y Arrio (1988), la bioluminiscencia en el caso de los dinoflagelados podría considerarse como un proceso metabólico involucrado en la regulación del exceso de poder reductor intracelular producido por la respiración y la fotosíntesis. Hoy en día, la hipótesis de la defensa contra los depredadores ha sido más investigada y establecida (Hanley y Widder, 2017), y la "defensa del oxígeno" (Wilson y Hastings, 2013) explicaría cómo surgió y evolucionó la bioluminiscencia antes de que desarrollara un efecto en otro grupo de organismos que podrían ejercer presión selectiva. Sin embargo, esto no implica que la bioluminiscencia tenga el mismo papel en el océano moderno. Los estudios integrados de regulación y evolución génica, combinados con factores ambientales e interacciones ecológicas, nos proporcionarían nuevas perspectivas para comprender el propósito de esta extraordinaria innovación funcional en dinoflagelados (Valiadi e Iglesias, 2013).

1.4 RETOS EN EL ESTUDIO PROTEÓMICO DE DINOFLAGELADOS

Los dinoflagelados poseen algunos de los genomas nucleares más grandes conocidos entre los eucariotas, que pueden ir desde 1.500 hasta 24.500.000 megabases (Mbp), por lo que ha sido difícil secuenciar el genoma completo de los miembros de este linaje (Wisecaver y Hackett, 2011) y este hecho ha constituido un obstáculo a la hora de llevar a cabo estudios

proteómicos en estos organismos. Características inusuales como la presencia de cromosomas permanentemente condensados a lo largo del ciclo celular (Herzog *et al.*, 1984; Dodge, 1985; Sigee, 1986) y minicírculos no genéticos en genomas de cloroplastos (Zhang *et al.*, 1999; 2002) han llevado al reconocimiento de los dinoflagelados como un solo grupo, Dinokaryote (Rizzo 1991; 2003; Soyer-Gobillard, 1996; Costas y Goyanes, 2005; De-Moreno-Díaz De-La-Espina *et al.*, 2005). Una de las características más destacadas de los dinoflagelados es su enorme y ampliamente variado genoma, alrededor de 3-245 gigabytes (Gb) de ADN, comprimido en algunos casos en más de 100 cromosomas, lo que representa de 1 a 80 veces más que un genoma haploide humano. Esto contrasta con la mayoría de las algas eucariotas, cuyo tamaño de genoma suele oscilar entre decenas y cientos de millones de pb (Mb) (Veldhuis *et al.*, 1997).

Sin embargo, en la última década se ha experimentado un rápido crecimiento en los estudios moleculares de estos organismos, especialmente en análisis transcriptómicos. Sobre la base de estos conjuntos de datos, se ha establecido que: en algunos casos más del 50 % de los genes no concuerdan con ningún gen documentado en ningún organismo; los genes dinoflagelados tienen más del 50 % de contenido de GC; solo alrededor del 10-27 % de los genes están regulados transcripcionalmente, algunos de los cuales muestran el ritmo circadiano; varios genes se expresan comúnmente y en un alto nivel en múltiples especies de dinoflagelados; hay una transferencia de genes desde el plástido al núcleo; los transcriptos del gen del plastidio están poli-uridiniladas en el extremo 3'; los ARNm nucleares sufren modificaciones (Splice-Lead (SL), trans-spliced) en el extremo 5'; los dinoflagelados tienen histonas nucleosómicas y proteínas relacionadas, y un conjunto de 79 proteínas ribosómicas eucarióticas. También se ha estudiado un pequeño número de genes a nivel genómico. La mayoría de estos genes existen en un gran número de copias dispuestas en repeticiones en tándem (Lidie y van Dolah, 2007; Zhang *et al.*, 2007; Bachvaroff y Place, 2008; Lin, 2011). La transcripción es de interés en el contexto de la estructura de la cromatina en los dinoflagelados, ya que muestra muchas diferencias con las células eucariotas más típicas. Los dinoflagelados muestran una clara escasez de factores de transcripción

específicos, aunque, sorprendentemente, el resto de la maquinaria básica de transcripción no es muy diferente de la encontrada en organismos filogenéticamente cercanos a los dinoflagelados (Roy y Morse, 2013).

1.4.1 Aplicación de tecnologías moleculares de alto rendimiento: transcriptómica y proteómica

La secuenciación masiva de ácidos nucleicos, como la transcriptómica, permite el análisis funcional de la biología microbiana en su conjunto. Los estudios de análisis basados en secuenciación de ácidos nucleicos o proteínas suelen utilizar un enfoque comparativo, y eso requiere secuencias de referencia relevantes, que son un recurso extremadamente limitado cuando se trata de eucariotas microbianos no modelo (Worden y Allen, 2010). En la práctica, esto significa que las bases de datos de secuencias deben estar pobladas con enormes cantidades de datos para los cuales tenemos certezas sobre la fuente. La más importante es la identidad taxonómica del organismo del cual se deriva una secuencia y la identificación funcional más alta posible de las proteínas codificadas. En un mundo ideal, dicha información estaría disponible como un gran conjunto de genomas completos, bien seleccionados y anotados. Sin embargo, la realidad se aparta sustancialmente de este ideal. Al menos para los eucariotas considerados no modelo, para los cuales el número de genomas disponibles es limitado y se ha basado en el crecimiento aleatorio de las bases de datos (Burki y Keeling, 2014; Keeling *et al.*, 2014). En comparación con los procariotas, los genomas nucleares eucarióticos son grandes y desproporcionadamente difíciles de analizar, lo que implica una baja cantidad de información molecular en las bases de datos. Esto da lugar a sesgos taxonómicos agudos en los genomas nucleares elegidos para la secuenciación, ya que una gran parte de ellos se derivan de organismos de particular importancia biomédica o biotecnológica.

Específicamente, la gran mayoría de los genomas nucleares provienen de animales, hongos y plantas, y de parásitos que infectan a los animales (Burki y Keeling, 2014; Del Campo *et al.*, 2014). Para los sistemas marinos, esto constituye una base de datos de referencia débil, ya que estos organismos son

colectivamente una representación limitada de la vida eucariótica en los mares. De hecho, los organismos marinos que mantienen la atmósfera terrestre, alimentan a las pesquerías del mundo y mantienen el ciclo global de carbono histórico (pre-antropogénico), así como los principales ciclos biogeoquímicos y de nutrientes en el océano, lamentablemente están fuera de estos grupos. Cada genoma secuenciado de un eucariota unicelular acuático ha proporcionado un conjunto de ideas nuevas e inesperadas (Armbrust *et al.*, 2004; Derelle *et al.*, 2006; Palenik *et al.*, 2007; Bowler *et al.*, 2008; Worden *et al.*, 2009; Gobler *et al.*, 2011; Nichols *et al.*, 2012; Curtis *et al.*, 2012; Read *et al.*, 2013; Shoguchi *et al.*, 2013). Sin embargo, debido a que los genomas nucleares pueden ser difíciles de secuenciar y ensamblar, y el modelado genético no siempre es sencillo, las necesidades inmediatas requieren una forma alternativa de generar una base de datos de referencia, siendo la transcriptómica la principal (Worden y Allen, 2010).

La secuenciación a gran escala del ARNm permite la caracterización rápida y eficiente de los genes expresados sin gastar recursos de secuenciación en las grandes regiones inter-génicas, intrones y ADN repetitivo tan comunes en los eucariotas, mientras que al mismo tiempo se eliminan muchos problemas de ensamblaje y la predicción y modelación de genes. Como primer paso, los datos transcriptómicos obtenidos a partir de cultivos puros son bloques de construcción adecuados para comenzar a ensamblar bases de datos de referencia para biología microbiana eucariótica. Este enfoque genera un gran número de secuencias de codificación, en forma de “*contigs*” ensamblados, de un organismo conocido (Keeling *et al.*, 2014). Se ha logrado un progreso significativo en el estudio transcriptómico en dinoflagelados, alimentado en gran parte por la reciente disponibilidad de secuenciación de bajo costo. La mayoría de los factores esperados en la maquinaria de transcripción están presentes en los dinoflagelados, al menos con respecto a sus homólogos entre los *Alveolata*. La excepción a esta regla general es que los factores de transcripción específicos parecen estar muy reducidos en cantidad y tipo en los dinoflagelados. Por lo tanto, si bien la transcripción general no se aparta de lo que se esperaba en el caso de los eucariotas, la orientación específica de los genes para el control transcripcional

puede diferir como resultado de la organización inusual de la cromatina en esta clase de organismos (Roy y Morse, 2013). Los recientes avances en la tecnología de secuenciación masiva han permitido la generación de bases de datos transcriptómicas que en el pasado habrían sido extremadamente difíciles de lograr, no solo en términos de dinero sino también de tiempo invertido. Por lo tanto, el volumen de datos transcriptómicos producidos ha crecido rápidamente y ahora se puede expandir a un transcriptoma de todo el genoma (Lin, 2011). Es probable que este esfuerzo proporcione un gran salto en nuestra comprensión y estrategias de investigación.

Por otro lado, varios estudios proteómicos en al menos 18 especies de microalgas (Tabla 1) han arrojado luz sobre los tipos de proteínas que pueden producir estos genomas, lo que es realmente importante al conocer los complejos mecanismos que utilizan las microalgas para producir proteínas, incluidos los mecanismos de trans-empalme y post-traducciona, entre otros.

Tabla 1. Especies de microalgas con estudios proteómicos hasta la fecha.

ESPECIES	REFERENCIA
<i>Chlamydomonas reinhardtii</i>	Nguyen <i>et al.</i> , 2011; Schmollinger <i>et al.</i> , 2014; Li <i>et al.</i> , 2017; Roustan <i>et al.</i> , 2017
<i>Chlorella protothecoides</i>	Gao <i>et al.</i> , 2014
<i>C. saccharophila</i>	Krohn <i>et al.</i> , 2017
<i>C. sorokiniana</i>	Ma <i>et al.</i> , 2013
<i>C. vulgaris</i>	Guarnieri <i>et al.</i> , 2011
<i>Dunaliella salina</i>	Gu <i>et al.</i> , 2015; Emami <i>et al.</i> , 2015; Ben <i>et al.</i> , 2017
<i>Emiliania huxleyi</i>	Read <i>et al.</i> , 2013
<i>Ettlia oleoabundans</i>	Garibay <i>et al.</i> , 2017
<i>Haematococcus pluvialis</i>	Gu <i>et al.</i> , 2015
<i>Lobosphaera incisa</i>	Siegler <i>et al.</i> , 2017
<i>Nannochloropsis oculata</i>	Kim <i>et al.</i> , 2005
<i>Nannochloropsis gaditana</i>	Fernandez-Acero <i>et al.</i> , 2018
<i>Phaeodactylum tricornutum</i>	Yang <i>et al.</i> , 2014; Longworth <i>et al.</i> , 2016
<i>Poterioochromonas malhamensis</i>	Beisser <i>et al.</i> , 2017
<i>Prototheca zopfii</i>	Murugaiyan <i>et al.</i> , 2013; Irrgang <i>et al.</i> , 2015; Murugaiyan <i>et al.</i> , 2016; Murugaiyan <i>et al.</i> , 2017
<i>Scenedesmus quadricauda</i>	Krohn <i>et al.</i> , 2017
<i>Tisochrysis lutea</i>	Garnier <i>et al.</i> , 2014

El objetivo principal de la mayoría de ellos estaba orientado a mejorar la producción de biodiesel (pero también se han abordado otros aspectos relevantes de su biología). Estos estudios han utilizado diversas estrategias proteómicas, desde el análisis MALDI TOF/TOF hasta los enfoques cualitativos y cuantitativos LC-MS/MS (Weretilnyk *et al.*, 2001; Chan *et al.*, 2004).

La combinación sinérgica entre el análisis transcriptómico y el modelado del proteoma mediante espectrometría de masas (LC-MS/MS) permitirá obtener una gran cantidad de información que hasta la fecha no está disponible en el caso de *P. lunula*. Está claro que un análisis *omico* (transcriptómico-proteómico) proporcionara información que tiene implicaciones importantes para los enfoques de biología de sistemas (Ghazalpour *et al.*, 2011), y la experiencia ha demostrado que dichos datos pueden transformar nuestra comprensión de la biología básica, al mismo tiempo que abre la puerta a nuevas direcciones de investigación.

1.5 APLICACIONES BIOTECNOLÓGICAS DE LAS MICROALGAS

El uso de microalgas también tiene grandes aplicaciones biotecnológicas debido, entre otras cosas, al potencial comercial de muchas moléculas bio-activas que pueden extraerse de ellas (Fajardo *et al.*, 2019). Las microalgas se pueden utilizar para producir diferentes tipos de metabolitos, como proteínas, lípidos, carbohidratos, carotenoides o vitaminas como aditivos para piensos, suplementos nutricionales humanos, cosméticos y para la obtención de energía, ya que son más eficientes en la conversión de la luz solar que las plantas superiores (Carrasco *et al.*, 2018). En la industria de la acuicultura, son la principal fuente de nutrición para las etapas iniciales de muchas de las especies utilizadas para la producción. Los suplementos terapéuticos de microalgas representan un mercado importante, y compuestos como β -caroteno, astaxantina, ácido graso poliinsaturado (PUFA), ácido docosahexaenoico (DHA), ácido eicosapentaenoico (EPA), y polisacáridos, como *b*-glucano, muestran el mejor potencial comercial. Las cepas de *Chlorella*, *Arthrospira* (Spirulina) y *Dunaliella* han sido el centro de investigación para la producción de biocombustibles, ya que convierten la energía solar en

varias formas químicas (Kruse *et al.*, 2005; Chisti 2007; Schenk *et al.*, 2008; Priyadarshani y Rath 2012; Villarruel-López *et al.*, 2017; Gatamaneni *et al.*, 2018; Kumar *et al.*, 2018; Pierobon *et al.*, 2018; Velazquez-Lucio *et al.*, 2018).

Ejemplos de las actuales aplicaciones biotecnológicas de la bioluminiscencia de *P. lunula* son los bioensayos Lumitox (Lumitox Gulf LC, EE.UU.) y QwikLite™ 200 (Assure Controls, EE.UU.), dos kits diseñados para medir la toxicidad de contaminantes en ambientes marinos. Tras la exposición a sustancias tóxicas, la bioluminiscencia en esta especie se extingue hasta en un 40 % después de una exposición de 2 h (Craig *et al.*, 2003; Stauber *et al.*, 2008). QwikLite™ fue desarrollado originalmente por el Centro de Sistemas de Guerra Espacial y Naval de EE. UU. Esta prueba mide una disminución de la bioluminiscencia azul-verde a 480 nm después de la exposición a un tóxico. Esta herramienta puede ser un sustituto útil para las pruebas a largo plazo; por ejemplo, para determinar rápidamente la toxicidad por amoníaco en los efluentes. Estas pruebas no requieren cultivo de algas ni conocimientos ecotoxicológicos específicos, por lo que son adecuadas para la detección rápida de toxicidad en un gran número de muestras o como parte de una batería de pruebas de toxicidad. Desde un punto de vista ecológico, existe la necesidad de ensayos rápidos con pruebas de punto final que puedan usarse como sustitutos de las pruebas estándar, más lentas. Las microalgas son indicadores sensibles del cambio ambiental y, como principales productores primarios que constituyen la fuente de alimento basal en las redes alimentarias marinas y de agua dulce, se utilizan en la evaluación de riesgos para evaluar los impactos de los contaminantes en los ecosistemas acuáticos (Stauber *et al.*, 2008).

Recientemente, Hildenbrand *et al.* (2015) presentaron un bioensayo toxicológico para evaluar rápidamente los casos potenciales de contaminación del agua subterránea asociada con la extracción de gas natural (actividad de perforación no convencional). Como en los otros casos mencionados anteriormente, la bioluminiscencia de *P. lunula* se puede cuantificar utilizando la espectrofotometría como una medida de la viabilidad del organismo, con una disminución de la producción bioluminiscente normal en relación a la exposición de concentraciones crecientes de sustancias tóxicas acuosas.

Nuevas herramientas derivadas de la comprensión de las reacciones bioluminiscentes han permitido el desarrollo de innumerables aplicaciones en biotecnología e ingeniería bioquímica. De este modo, se ha desarrollado una amplia gama de técnicas analíticas *in vitro* e *in vivo* basadas en diferentes sistemas bioluminiscentes, que incluyen pruebas para diferentes analitos, inmunoensayos, ensayos de expresión génica, detección de drogas, bioimagen de organismos vivos, estudios sobre cáncer, investigación de enfermedades infecciosas y monitoreo ambiental (Kaskova *et al.*, 2016). Con la reciente invención de los reporteros luminiscentes, las posibilidades futuras para el desarrollo de aplicaciones de reporteros adicionales son prometedoras (Sharifiana *et al.*, 2017). Por ejemplo, para ensayos de reportero doble y múltiple, es posible intercambiar LCF de dinoflagelados por LCFs de luciérnaga o de cnidario (*Renilla reniformis*), y usarla para genes de control o “reporteros”. Por lo tanto, un ensayo de reportero triple puede identificar tres actividades transcripcionales de diferentes genes al mismo tiempo (Suzuki *et al.*, 2005). Además, estas capacidades pueden mejorarse en gran medida mediante las adaptaciones de enzimas y sustratos nativos, que se originan a partir de organismos luminiscentes de diversos linajes evolutivos. Ha sido posible diseñar tanto una enzima como un sustrato en un nuevo sistema de bioluminiscencia capaz de emitir luz de manera más eficiente y con características bioquímicas y físicas superiores (Hall *et al.*, 2012).

Por otro lado, y en función de la estimulación mecánica en el entorno natural, se ha explorado el potencial de que la bioluminiscencia emitida podría utilizarse para localizar y rastrear las perturbaciones en el agua producidas por nadadores e incluso objetivos militares (Moline *et al.*, 2007; Marcinko *et al.*, 2013). Otra aplicación de la bioluminiscencia de los dinoflagelados la ilustra un nuevo enfoque para pronosticar el potencial de bioluminiscencia a través del modelado de la dinámica poblacional de este género. El modelo desarrollado ofrece una plataforma prometedora para el futuro pronóstico operacional de los amplios cambios temporales en la bioluminiscencia en el Atlántico Norte. Dicha previsión de patrones estacionales podría proporcionar información valiosa para la orientación de campañas científicas de campo (Marcinko *et al.*, 2013; 2014). Tokarev *et al.* (2017) señalaron además que sobre la base de datos

obtenidos desde el espacio sobre de la distribución espaciotemporal del fitoplancton, acoplados a las mediciones de bioluminiscencia *in situ*, es posible crear algoritmos regionales para la transición desde las estimaciones numéricas del fitoplancton (clorofila a) al potencial bioluminiscente en el día y a la proyección de modelos inversos durante la noche. Las tendencias de la bioluminiscencia en la ciencia se desarrollan muy rápidamente y las ganancias globales de las tecnologías de bioluminiscencia se estiman en aproximadamente 5 mil millones de dólares por año (Kaskova *et al.*, 2016).

1.6 INGENIERÍA GENÉTICA EN MICROALGAS

En los últimos 10 años, se han logrado avances significativos en el campo de la manipulación genética en microalgas. Esto incluye la expresión eficiente de transgenes utilizando el nuevo mecanismo de ribo-switches para la regulación de genes (Croft *et al.*, 2007), el acoplamiento de promotores nucleares y el uso de la LCF como reportero para la expresión de genes insertados (Shao y Bock 2008), y el uso de cloroplastos inducibles para la expresión génica (Surzycki *et al.*, 2007; Bogen *et al.*, 2013). El uso de microalgas en este campo presenta varias ventajas en comparación con otras plataformas basadas en mamíferos ya que el período de tiempo desde la transformación temprana hasta el nivel de producción se reduce considerablemente (Blatti *et al.*, 2013). Las tecnologías actuales han permitido la manipulación de uno o más genes para estudiar la función basada en el sistema y otras aplicaciones moleculares. Las aplicaciones recientes de la edición de genes se pueden utilizar como una plataforma de diseño novedoso para proteínas y modelado computacional para la mejora de la producción de diferentes moléculas (Cock *et al.*, 2010; Gupta y Shukla, 2015; Shukla y Karthik, 2015; Banerjee *et al.*, 2016; Anand *et al.*, 2017).

Sin embargo, en los últimos años muchas de las conclusiones biológicas obtenidas mediante biología molecular han requerido la generación de mutantes. Esta generación eficiente de mutantes es importante para (i) análisis funcionales de genes/proteínas para mejorar nuestra comprensión de la biología de las microalgas y (ii) para la mejora de las cepas para aplicaciones

biotecnológicas. Hasta hace poco, solo ha sido posible el silenciamiento de la expresión génica (ARNi) o el uso de líneas con genes sobre-expresados en diatomeas. Hoy en día, se ha demostrado que la edición dirigida del genoma es posible, pero solo para unas pocas especies de microalgas. Algunos ejemplos de esto pueden ser: i) el empleo de recombinación homóloga para la eliminación de nitrato y nitrito reductasa y ferritina en *Nannochloropsis sp.* y *Ostreococcus sp.* (Kilian *et al.*, 2011; Lozano *et al.*, 2014); ii) el uso de nucleasas de dedo de zinc para la inserción de marcadores de selección y la modificación de genes nativos en *Chlamydomonas reinhardtii* (Sizova *et al.*, 2013); iii) el uso de nucleasas efectoras de tipo activador transcripcional (TALEN) para la mutagénesis dirigida, o iv) la inserción de genes en *Phaeodactylum tricornutum* (Daboussi *et al.*, 2014; Weyman *et al.*, 2015). Como un ejemplo más reciente tenemos el uso del sistema de repetición palindrómica corta (CRISPR)/Cas9 agrupado en intervalos regulares interpolados para la introducción de mutaciones estables de genes dirigidos en *C. reinhardtii* (Shin *et al.*, 2016) y *P. tricornutum* (Nymark *et al.*, 2016).

En este punto, resulta muy interesante considerar el enfoque propuesto por el campo de la biología sintética, es decir, cambiar el enfoque original de los sistemas biológicos para nuevos objetivos y aplicaciones. A través de la expresión coordinada y equilibrada de genes, tanto nativos como introducidos desde otros organismos, los recursos pueden ser desviados para la producción comercial de productos de alto valor. Este campo interdisciplinario en desarrollo tiene el potencial de revolucionar el descubrimiento de productos naturales a partir de plantas superiores, al proporcionar una amplia gama de herramientas, tecnologías y estrategias para explorar la gran variedad de productos naturales que producen los organismos unicelulares (Moses *et al.*, 2017; Jagadevan *et al.*, 2018). Las microalgas son plataformas de biotecnología industrial adecuadas, ya que, como las bacterias y las levaduras, se pueden cultivar en gran escala en entornos controlados y contenidos. Sin embargo, al ser fotosintéticos y poder fijar carbono de fuentes inorgánicas, su cultivo es en principio más sostenible que el de los sistemas heterótrofos; aunque es esencial llevar a cabo una evaluación del ciclo de vida para iniciar cualquier proceso de diseño en particular (Kazamia y Smith 2014). Se deben superar

muchos desafíos antes del cultivo de cepas de ingeniería a escalas muy grandes necesarias para una producción económicamente rentable, (Scott *et al.*, 2010; Klein-Marcuschamer *et al.*, 2013; Scaife y Smith, 2016). Si bien estos desafíos se pueden abordar con las actuales herramientas genéticas y moleculares, la adopción de los principios de biología sintética puede acelerar el engorroso proceso de desarrollo y optimización de cepas (Moses *et al.*, 2017). Para este fin, las tecnologías de ingeniería actuales de las células fotosintéticas deben actualizarse y, finalmente, se deben desarrollar nuevas herramientas para producir productos bioquímicos de algas económicamente competitivos (Ng *et al.*, 2017). El enfoque metabólico o el de la biología sintética ha recibido la mayor atención, ya que el diseño de una vía metabólica en un nuevo sistema abre posibilidades para la industrialización de las microalgas (Andrianantoandro *et al.*, 2006). En biología sintética, el proceso básico de manipulación de genes sigue siendo el mismo, incluida la selección de hospedadores, genes de diana, construcción de plásmidos, herramientas de transformación, sistema de selección y herramientas de edición de ADN (Ng *et al.*, 2017; Jagadevan *et al.*, 2018).

La limitada información genómica disponible hasta la fecha en el caso de *P. lunula* es un obstáculo importante para llevar a cabo estudios a nivel molecular en esta especie. Impulsado por este desafío, nuestro principal objetivo ha sido generar un transcriptoma y proteoma de referencia de *P. lunula* que apoyará y complementará futuras líneas de investigación en esta especie. Todos los datos generados en este estudio servirían como un punto de partida y/o marco de referencia para la eventual aplicación de técnicas de ingeniería genética con el fin de desarrollar nuevas aplicaciones biotecnológicas y/o mejorar las ya existentes.

1.7 HIPÓTESIS Y OBJETIVOS

La hipótesis de partida de la presente Tesis Doctoral se basa en que dada la naturaleza bioluminiscente de *P. lunula* y su complejidad genómica, es necesario profundizar en la biología y en la expresión proteica de este dinoflagelado, haciendo posible la identificación de una o más proteínas derivadas de la determinación del proteoma de esta especie que puedan ser de interés desde el punto de vista de la biología básica o para el sector biotecnológico y/o agroindustrial.

De este modo, en la presente Tesis Doctoral se ha plantado el siguiente **Objetivo General**: realizar una aproximación a la biología de *P. lunula*, mediante técnicas de alto rendimiento (transcriptómica y proteómica), para la caracterización del proteoma de *P. lunula* con el objeto de buscar proteínas de interés para el sector biotecnológico y/o agroindustrial.

Para lograr el citado Objetivo General se han planteado los siguientes **Objetivos Específicos**:

- Optimización del protocolo de extracción de ARN y de proteínas totales.
- Análisis de secuenciación masiva del ARNm.
- Análisis bioinformático para el ensamblaje del transcriptoma de referencia.
- Obtención de los perfiles proteicos mediante LC-MS/MS.
- Análisis bioinformático de las identificaciones derivadas de la determinación del proteoma: ontología genética (GO) e identificación de códigos enzimáticos (KEGG).
- Determinación de proteínas de interés.
- Establecimiento de perspectivas futuras de investigación.

2. MATERIALES Y MÉTODOS

Debido a la falta de información genómica sobre *P. lunula*, se generó un transcriptoma de referencia mediante secuenciación masiva y cuyo ensamblaje se realizó utilizando los tres principales programas bioinformáticos actualmente utilizados para este propósito (RAY, Trinity y MIRA). Paralelamente, se llevó a cabo la purificación de extractos de proteínas totales, posteriormente analizados por LC-MS. Luego, para evaluar la calidad de los diferentes ensamblajes obtenidos, se realizó un análisis proteómico múltiple, utilizando los proteomas inferidos de todos los transcriptomas ensamblados de *P. lunula* como base de datos de búsqueda. En consecuencia, el ensamblaje más adecuado fue seleccionado y refinado para producir finalmente un proteoma de referencia completo de *P. lunula* de alta calidad. Se realizó un análisis proteómico afinado contra este último, dando información muy relevante sobre la composición del proteoma de *P. lunula*. Además, se realizó un análisis funcional de las proteínas identificadas basado en términos de ontología génica (GO), códigos de enzimas y asociaciones de rutas (KEGG), para obtener una imagen amplia de las funciones celulares representadas por los componentes moleculares aquí analizados. Finalmente, se llevó a cabo un proceso de curación de información y extracción de datos para resaltar algunos de los componentes moleculares más relevantes identificados dentro del proteoma de *P. lunula*.

2.1 CONDICIONES DE CULTIVO

Las muestras de *P. lunula* (CCAP 1131/1) se cultivaron en medio F2 (Guillard y Ryther, 1962) en un fotoperíodo de 12 h de oscuridad y 12 h de luz a una temperatura de 18-20 °C y una salinidad de 31 g / l. Las muestras se recogieron al comienzo de la fase oscura.

2.2 AISLAMIENTO DE ARN TOTAL

El ARN se extrajo usando el kit NucleoSpin RNA® (Macherey-Nagel), siguiendo las indicaciones del fabricante, y se cuantificó por espectrofotometría UV, así como por electroforesis en geles de agarosa. La calidad del ARN aislado se evaluó utilizando el sistema 2100 Bioanalyzer (Agilent Technologies),

con el kit RNA 6000 Nano®, siguiendo las indicaciones del fabricante. Solo se utilizó ARN de alta calidad (RIN = 8.3 y relación A260: A280 cerca de 2.0) para experimentos posteriores.

2.3 SECUENCIACIÓN DE TRANSCRIPTOMA

La biblioteca de ADNc se construyó utilizando 2 ug de ARN total, en un Sistema Ion Chef (Thermo Fisher Scientific) y el Kit Ion Total RNA-Seq v2 para bibliotecas de transcriptomas completos (Life Technologies Corporation, California, EE. UU.), de acuerdo con las instrucciones del fabricante. La secuenciación de próxima generación (NGS) se llevó a cabo utilizando la tecnología de semiconductores a través de un sistema Ion S5 (Thermo Fisher Scientific) utilizando un chip de secuenciación Ion 540.

2.4 ENSAMBLAJE *DE NOVO* Y ANÁLISIS DE SECUENCIACIÓN DE ARN DE ALTO RENDIMIENTO

Las lecturas sin procesar obtenidas de la plataforma de secuenciación se preprocesaron para retener solo secuencias de alta calidad para ser utilizadas posteriormente para los ensamblajes (Cutadapt versión 1.9, BBDuk versión 35.43). Los adaptadores de secuencia se recortaron primero, y las bases de baja calidad (con puntaje de phred por debajo de un umbral) se recortaron en secuencias sin procesar. Se seleccionaron dos valores diferentes de puntuación de phred como umbrales (15 y 20), produciendo dos conjuntos de datos de secuencia. Luego, estos conjuntos de datos se procesaron en dos, cada uno filtrando lecturas de menos de 30 o 100 nt, para un total de cuatro conjuntos de datos de secuencia etiquetados de acuerdo con los filtros de calidad y longitud que pasaron (Q15 L30, Q15 L100, Q20 L30 y Q20 L100) La evaluación de la calidad de las lecturas se realizó con el software FastQC (versión 0.11.3) (Wingett y Andrews, 2018) para evaluar el efecto de cada paso de este proceso. Todos los análisis posteriores con el software RAY se llevaron a cabo utilizando estos cuatro conjuntos de datos de alta calidad para evaluar el efecto del tamaño de la biblioteca y el umbral de calidad en el rendimiento del algoritmo de ensamblaje probado. El ARN ribosómico restante fue

detectado por el software SortMeRNA (versión 2.0) (Kopylova *et al.*, 2012). Las lecturas recortadas de los adaptadores según la base de datos preempaquetadas de SortMeRNA id98 (de Silva v.119 y Rfam) con los parámetros predeterminados.

Se emplearon tres softwares diferentes para ensamblar el transcriptoma de *P. lunula* a través de un enfoque *de novo*, ya que no hay un genoma de referencia disponible hasta la fecha, y se evaluó adicionalmente para contrastar los ensamblajes obtenidos mediante un análisis proteómico. Los cuatro conjuntos de datos de secuencia preprocesados descritos anteriormente se ensamblaron en “contigs” usando 5 parametrizaciones diferentes de kmer (21, 31, 41, 51, 71) con el ensamblador RAY 2.3.1. (Boisvert *et al.*, 2010), para un total de 20 ensamblajes. RAY se usó para ensamblar pares de lectura que se superponen inequívocamente, así como los extremos emparejados no fusionados como un grupo, ya que utiliza gráficos de Bruijn que emplean puntos definidos que se pueden extender con el algoritmo. El proceso de extensión se lleva a cabo con un enfoque heurístico que puede detenerse si la familia de lecturas no indica claramente la dirección de la extensión (Boisvert *et al.*, 2010). Además, para generar más robustez en los ensamblajes, debido a la falta de un genoma de referencia, el software Trinity y Mira también se utilizaron para generar ensamblajes *de novo*.

Debido a que los resultados del ensamblador RAY fueron mejores con el conjunto de datos de secuencia Q20 L100, y a los enormes recursos computacionales requeridos por Trinity y Mira con el aumento del archivo de lecturas de entrada, solo este conjunto de datos de secuencia se empleó con los ensambladores Trinity y MIRA. Trinity 2.5.1, que realiza un ensamblaje *de novo* utilizando un algoritmo basado también en gráficos de Bruijn (Grabherr *et al.*, 2011), se utilizó con los siguientes parámetros: “—max_memory 1000G—CPU 20 --SS_lib_type F—bflyCalculateCPU—normalize_max_read_cov 20”, con “—KMER_SIZE 25” or “31”, también con y sin “—min_kmer_cov 2”, produciendo un total de 4 ensamblajes. El software MIRA 4.9.6 (Chevreux *et al.*, 1999), a diferencia de Trinity y RAY, se basa en la estrategia conocida como Overlap / Layout / Consensus, y se empleó para producir un ensamblaje con la

colección de secuencias Q20L100 bajo los siguientes parámetros: “job = denovo,est,accurate; COMMON_SETTINGS - GENERAL:number_of_threads=12 - KMERSTATISTICS:lossless_digital_normalisation=yes; IONTOR_SETTINGS - ALIGN:min_relative_score=70 -ASSEMBLY:minimum_read_length=100; - CLIPPING:quality_clip=no -CLIPPING:qc_minimum_quality=15 - CLIPPING:qc_window_length=20; -CLIPPING:clip_polyat=yes - CLIPPING:cp_min_sequence_len=12; technology = iontor”. La evaluación de la estructura de los 25 conjuntos generados se realizó mediante un script R interno. Además, se realizó una búsqueda BLAST para comparar todos los ensamblajes entre sí (NCBI BLAST 2.2.31+) (Camacho *et al.*, 2009) para evaluar su representatividad.

2.5 ANÁLISIS PROTEÓMICO

La lisis mecánica de las muestras se realizó con un homogeneizador Minilys® (Bertin Technologies). La precipitación de proteínas se llevó a cabo utilizando un procedimiento a base de TCA/acetona descrito previamente (Liñeiro *et al.*, 2016). El ensayo se realizó por triplicado. Una vez que se obtuvieron los extractos de proteína de *P. lunula*, se estimó la concentración de proteína por medición fluorométrica usando Qubit System® (Thermo Fisher Scientific).

La digestión de proteínas se llevó a cabo con tripsina en 25 mM de bicarbonato de amonio a 37 °C durante 3 h. Anteriormente, el extracto de las proteínas fue limpiado mediante 1D SDS-PAGE, concentrando el extracto total en 1 cm de gel de resolución al 10%. Las proteínas se visualizaron con azul brillante Coomassie G-250. Luego, las láminas de gel se cortaron en cubitos y se tiñeron con 50% de acetonitrilo en 100 mM de bicarbonato de amonio. Para bloquear los grupos cisteína tiol, se aplicaron 25 mM de ditiotretol y 40 mM de yodoacetamida en 25 mM de bicarbonato de amonio a las láminas de gel.

La separación de los péptidos mediante nano-LC se realizó en un Dionex Ultimate 3000 nano UPLC (Thermo Fisher Scientific) con un i.d. de 75

µm. x columna Acclaim PepMap RSLC C18 de 50 cm (Thermo Fisher Scientific) a 300 nl / min y 40 ° C. La mezcla de péptidos fue previamente desalada y limpiada en un 300 µm i.d. x cartucho Acclaim PepMap de 5 mm (Thermo Fisher Scientific) en acetonitrilo al 2% / ácido fórmico al 0,05% durante 3 min a 5 µl / min. El tampón A (ácido fórmico al 0,1%) y el tampón B (acetonitrilo al 80%, ácido fórmico al 0,1%) se utilizaron como fase móvil en separaciones cromatográficas con las siguientes condiciones: 4-35% de B durante 60 min; 35-55% de B durante 3 min; 55-90% de B durante 3 minutos, seguido de 8 minutos de lavado al 90% de B y 12 minutos de reequilibrio al 4% de B. Los péptidos positivos de elución se analizaron en modo positivo en un espectrómetro de masas Orbitrap Fusion™ Tribrid™ (Thermo Fisher Scientific) con una fuente de ionización por nanoelectrospray que funciona en modo de adquisición de datos dependientes Top30 con un tiempo de ciclo máximo de 3 s. Se obtuvieron exploraciones completas de MS1 en un rango de masa de 400 a 1500 m / z a una resolución de 120,000 (a 200 m / z). Solo aquellos iones precursores con estado de carga 2-5 fueron muestreados para MS / MS. Los iones monoisotópicos, previamente aislados en el cuadrupolo a 1,2 Da, se fragmentaron con CID en la trampa de iones con una energía de colisión normalizada al 35%. El control automático de ganancia fue de 2¹⁰3 y el tiempo máximo de inyección fue de 300 ms. Se estableció un tiempo de exclusión dinámica en 15 s con una tolerancia de 10 ppm alrededor del precursor seleccionado y sus isótopos. Estos estudios se realizaron en colaboración con el SCAI de la Universidad de Córdoba.

2.6 ONTOLOGÍA GENÉTICA Y ANÁLISIS KEGG

La anotación funcional de los términos de ontología génica (GO), códigos de enzimas, asociaciones de vías InterPro y KEGG se realizó utilizando el programa BLAST2GO (Conesa *et al.*, 2005; Conesa y Götz, 2008; Olaya-Abril *et al.*, 2018). Los códigos de enzimas asociadas se mapearon para hacer referencia a las rutas KEGG (KO) para ilustrar mejor los componentes representativos de la ruta identificados en *P. lunula* (herramienta de mapeo KEGG).

2.7 VERIFICACIÓN DE LA PRESENCIA DE LBP Y LCF / LBP POR PCR

La purificación del ARN total se realizó como se describió anteriormente. La síntesis de ADNc y la amplificación de los fragmentos por RT-PCR se llevó a cabo utilizando el kit qScript® (Quanta Biosciences), Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific) y Q5® High-Fidelity DNA Polymerase (New England Biolabs) Inc.), siguiendo las recomendaciones de los fabricantes. Las condiciones de la reacción para amplificar LBP fueron las siguientes: 2X Phusion Flash PCR Master Mix 10 µL, 10 µM LBP F: 5'-TGT GGC GGA CCA TGG TGG A-3' 1 µL, 10 µM LBP R 5'-GGT CTT TGA TGG AGA GCT CCA-3' 1 µL, ADNc (65 ng/µL) 1 µL y agua 7 µL, para un volumen total de reacción de 20 µL.

Con los cebadores LBP F y LBP R se obtuvo un fragmento de 453 pb. El programa del termociclador fue el siguiente: desnaturalización inicial a 98 °C durante 10 segundos, seguido de 35 ciclos a 98 °C durante 1 segundo, 66 °C durante 5 segundos, 72 °C durante 8 segundos, con un ciclo de extensión final a 72 °C durante 1 minuto y parada 4 °C. En el caso de la isoforma híbrida LCF / LBP, las condiciones fueron las siguientes: 5X Q5 tampón de reacción 5 µL, 10 µM LCF/LBP F: 5'-GAG GAC TAC TCG AAC GGA TTG ATC C-3' 1.25 µL, 10 µM LCF/LBP R: 5'-GTC CAT GAT GCG CCA GGG CT-3' 1.25 µL, 10 mM dNTPs 0.5 µL, ADNc (65 ng/µL) 1 µL, Q5 High-Fidelity DNA polymerase 0.25 µL, y agua 15.75 µL, para un volumen total de reacción de 25 µL.

Con los cebadores LCF / LBP F y LCF / LBP R se obtuvo un fragmento de 1,395 pb. El programa del termociclador fue el siguiente: desnaturalización inicial a 98 °C durante 30 segundos, seguido de 35 ciclos a 98 °C durante 10 segundos, 69 °C durante 20 segundos, 72 °C durante 45 segundos, con un ciclo de extensión final a 72 °C durante 2 minutos y parada a 4 °C. Para la purificación de los fragmentos LBP y LCF / LBP, se siguieron las instrucciones del kit de extracción de gel GeneJET (Thermo Fisher Scientific). Los fragmentos se visualizaron por electroforesis (agarosa al 1%, teñida con gel red) durante 1 hora a 110 V. Los resultados se corroboraron mediante secuenciación de Sanger y análisis BLASTx (Altschul *et al.*, 1997).

2.8 ANÁLISIS FILOGENÉTICOS Y ESTRUCTURALES

Las secuencias de las especies de dinoflagelados para los genes LCF, LBP y glutatión S-transferasa (GST) se descargaron de GenBank (Tabla 2). Las secuencias se alinearon con el software MUSCLE (Edgar, 2004) y el análisis filogenético se realizó con el método de máxima verosimilitud basado en el modelo de tiempo general reversible (Tamura y Nei, 1993) con el software MEGA v 7.0.14 (Kumar *et al.*, 2016) con 1000 valores de arranque (Nei y Kumar, 2000). Se utilizó una distribución Gamma discreta para modelar las diferencias de velocidad evolutiva entre los sitios (5 categorías). La predicción de la estructura 2D / 3D de las regiones conservadas analizadas se realizó utilizando el servidor web Phyre2 (Kelley *et al.*, 2015, [www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id = índice](http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index)) y las estructuras resultantes se visualizaron con el software de modelado de moléculas EzMol (versión 1.22, www.sbg.bio.ic.ac.uk/ezmol) (Reynolds *et al.*, 2018). El logotipo de secuencia se realizó utilizando WebLogo (Crooks *et al.*, 2004, weblogo.berkeley.edu).

Tabla 2. Números de acceso de las secuencias reportadas en GenBank para la luciferasa (LCF), la proteína de unión a luciferina (LBP) y la glutatión S-transferasa (GST) en microalgas.

ESPECIES	LCF	LBP	GST
<i>Alexandrium affine</i>	AAV35377	AFN26992	
<i>Alexandrium catenella</i>		ABY78836	
<i>Alexandrium fundyense</i>	AEW67906	AFN26994	
<i>Alexandrium monilatum</i>	AEW67931	AFN26995	
<i>Alexandrium ostenfeldii</i>	AOG16037		
<i>Alexandrium tamarense</i>	AAV35378	AFN27008	
<i>Coratocorys horrida</i>	AEW67919	AFN27015	
<i>Gonyaulax spinifera</i>	ABO61069		
<i>Lingulodinium polyedra</i>	O77206	AAA29165, AAA29166	
<i>Noctiluca scintillans</i>	AED02505	AHB24369	
<i>Protoceratium reticulatum</i>	AAV35381	AFN27016	
<i>Pyrocystis fusiformis</i>	AAV35379		
<i>Pyrocystis lunula</i>	AAL40676, AAL40677, AAL40678	MN259726, MN259727	AAN85429
<i>Pyrocystis noctiluca</i>	AAV35380		
<i>Pyrodinium bahamense</i>	KX377172		
<i>Tripos digitatus</i>	AEW67915		

3. RESULTADOS

3.1 ENSAMBLAJE DE TRANSCRIPTOMA *DE NOVO*

Se generaron un total de 80,874,825 lecturas sin procesar (11,292,087,505 pb, 55.82% GC, lecturas de longitud media de 140 pb). El porcentaje de ARN ribosómico residual que queda en la muestra después de la etapa de enriquecimiento en ARN mensajero se cuantificó como se detalla en la sección de material y métodos. Los ARN ribosómicos identificados se clasificaron para identificar la posible contaminación por procariotas presentes en la muestra. Por lo tanto, del 11.39% de las lecturas que superan el umbral en la búsqueda de ARNr (9.212.736 lecturas), el 9.16% y el 1.87% coincidieron con los ribosomas eucariotas 18s y 28s, respectivamente, mientras que solo el 0.24% y el 0.12% coincidieron con los bacterianos 16s y 23s, respectivamente. Las lecturas sin procesar generadas para el ensamblaje *de novo* se depositaron en la base de datos NCBI SRA, con el número de acceso BioProject PRJNA497104.

Se generaron un total de 25 ensamblajes y se evaluaron adicionalmente para seleccionar el algoritmo y la parametrización más adecuados para un ensamblaje de transcriptoma de *P. lunula* exitoso. Así, se realizó un análisis de la estructura de ensamblaje, para contigs mayores de 400 nt. Atendiendo a los tres requisitos de cálculo de los algoritmos (Guerrero-Sánchez *et al.*, 2019), RAY se empleó primero para analizar el efecto de establecer el parámetro kmer, el tamaño de la biblioteca y la calidad de lectura (en términos de tamaño mínimo de lectura y puntajes medios de phred) en el rendimiento del algoritmo. Los resultados muestran claramente que aumentar el parámetro kmer tiene un efecto tremendamente nocivo en el número final de contigs obtenidos, aunque, como era de esperar, la distribución del tamaño de contigs mejora con valores de kmer más altos, especialmente mostrando Nx más alto para kmer sobre 41 (Fig. 5).

Sin embargo, se obtuvo una estructura casi idéntica para todos los conjuntos de datos cuando el parámetro kmer se estableció dentro de 21-41, lo que sugiere que valores de kmer más bajos (dentro de 21-31) podrían ser apropiados en este caso, evitando una pérdida importante de nuevos contigs.

De hecho, cuando cada combinación de kmer se explora en detalle para evaluar el efecto del tamaño de la biblioteca y de los filtros de calidad de las lecturas, se obtiene una estructura ligeramente mejor cuando se aplica el Q20 (filtro de la más alta calidad).

Con respecto a los filtros de longitud de las lecturas (L100 y L30), ya que ambos parecen comportarse de manera similar para la estructura del ensamblaje, decidimos seguir utilizando la biblioteca filtrada más estricta (L100), porque esta es realmente la colección de secuencias más segura (calidad más estricta y filtros de longitud aplicados). Además, su tamaño reducido se adapta mejor a los programas Trinity y Mira, ya que estos dos algoritmos presentan requisitos computacionales mucho más altos que Ray para grandes conjuntos de datos (Guerrero-Sánchez *et al.*, 2019). Finalmente, al comparar los 10 ensamblajes obtenidos de la biblioteca Q20 L100 (Fig. 6), el obtenido con el software Trinity con los parámetros k25_NC20, presentó el mejor equilibrio en términos de estructura y número total de contigs.

Luego, para evaluar la representatividad de cada ensamblaje, se realizó una búsqueda BLASTn para cada secuencia en cada ensamblaje contra cada secuencia en cada ensamblaje (búsqueda recíproca de similitud de todos contra todos), incluyendo dos conjuntos de datos adicionales descargados de los repositorios, una cepa CCCM 517 de *P. lunula* (SRA: SRS621557), y el otro de *Breviolum minutum* (anteriormente *Symbiodinium minutum*, SRA: DRS003180) (Shoguchi *et al.*, 2013). Se emplearon un total de 3.645 búsquedas BLASTn (conjuntos de 27x27 en 5 umbrales de valor e) para recuperar el mejor resultado pasando un umbral de valor e específico [Material Suplementario 1]. Se determinó el porcentaje de secuencias de un ensamblaje que encuentran un hit en los demás ensamblajes (Fig. 7).

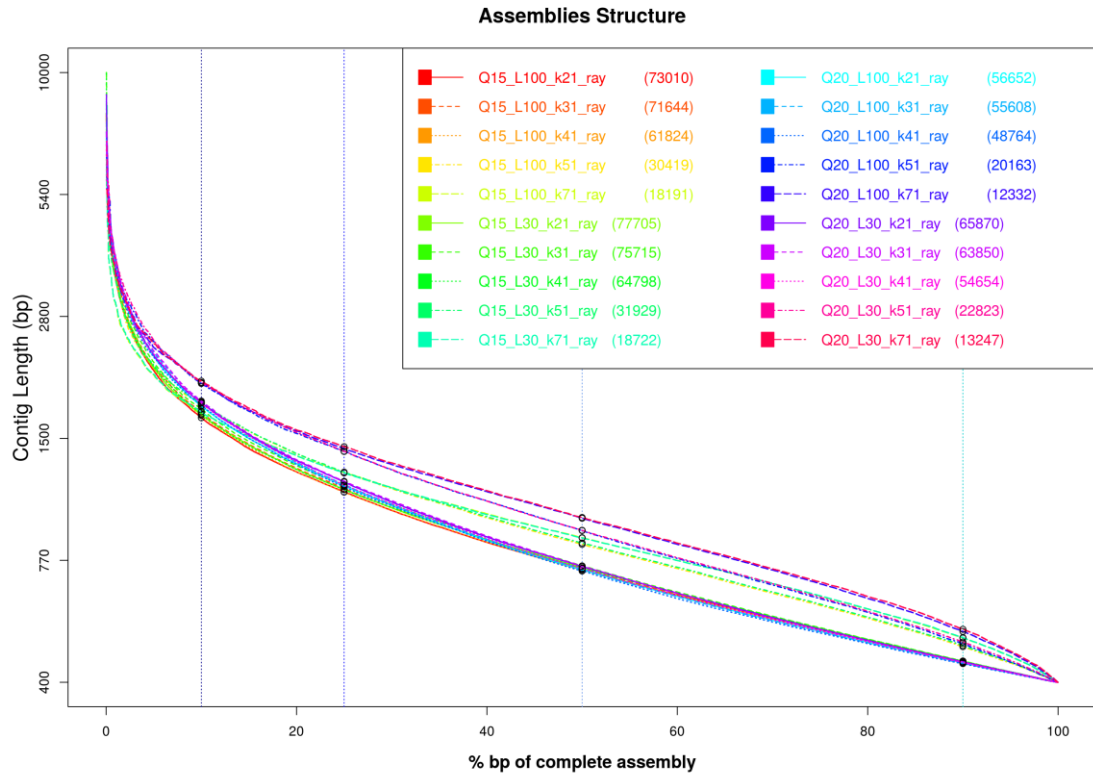


Figura 5. Comparación del análisis estructural de 20 ensambles transcriptómicos generados por RAY, utilizando diferentes filtros de puntaje phred (Q), longitud (L) y kmer (k).

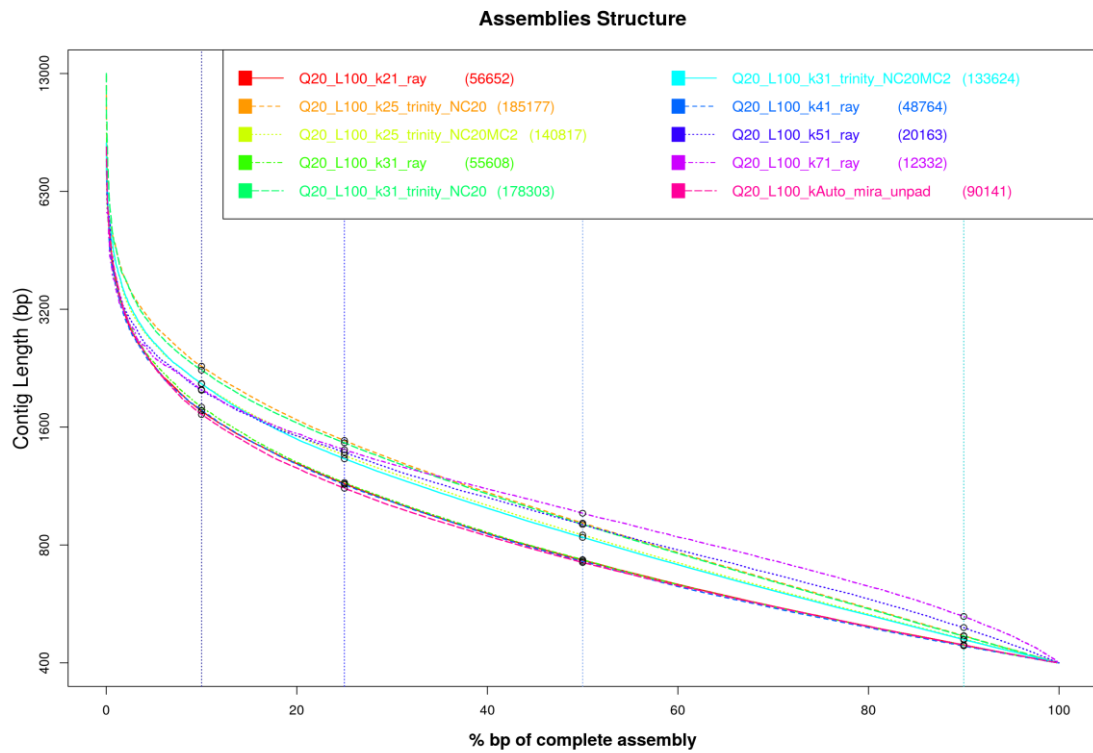


Figura 6. Comparación del análisis estructural de 10 ensambles transcriptómicos generados por los tres algoritmos (RAY, Trinity, MIRA) utilizando los conjuntos de datos con puntaje phred Q20 y longitud L100 con diferentes kmers (k).

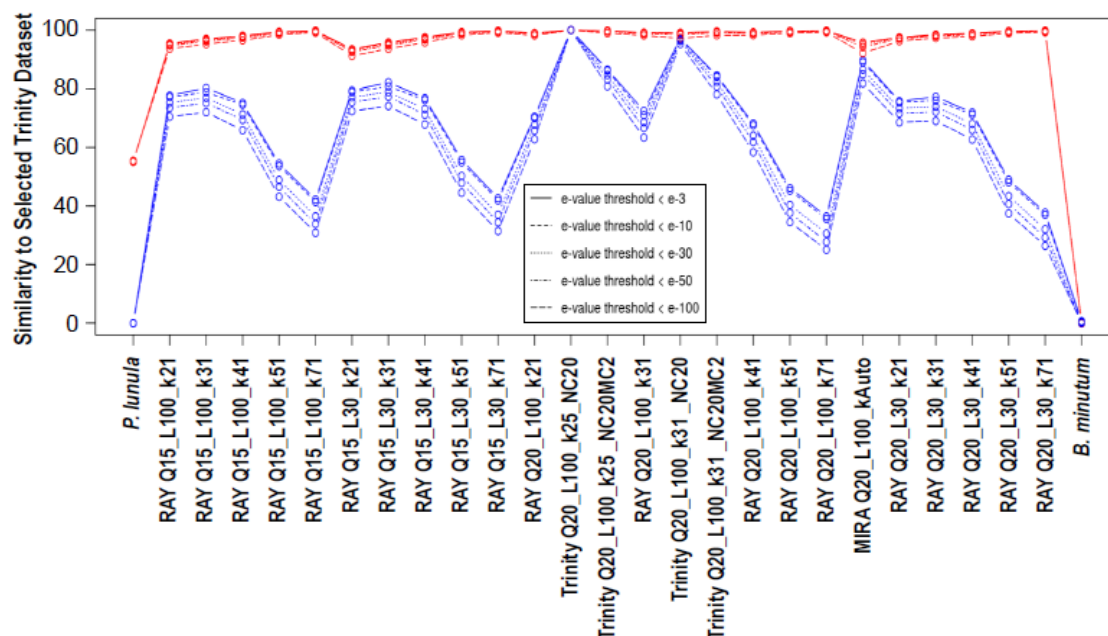


Figura 7. Análisis de representatividad por similitud de secuencia del conjunto Trinity_Q20 L100_k25_NC20 (líneas rojas), contra los 25 conjuntos, así como dos conjuntos de datos públicos de *P. lunula* (SRA: SRS621557) y *B. minutum* (SRA: DRS003180).

Para facilitar la visualización, solo se trazaron los resultados de las comparaciones recíprocas de Trinity_Q20 L100_k25_NC20 con los 27 ensamblajes. Los resultados indican claramente que, si bien casi todas las secuencias de la mayoría de los ensamblajes presentaron un hit, lo que significa que también están presentes en el ensamblaje evaluado por Trinity (líneas rojas), los otros ensamblajes claramente pierden muchos de los contigs que se presentan en el ensamblaje evaluado (líneas azules). Teniendo en cuenta los dos conjuntos de datos externos, los resultados manifiestan que el conjunto de datos de la cepa CCCM 517 de *P. lunula* es muy pobre (solo 9 secuencias más largas que 400 nt), y que la similitud en el nivel de nucleótidos de *P. lunula* con *B. minutum* es escasa.

La evaluación final de los ensamblajes generados fue realizada por proteómica. Por lo tanto, los proteomas inferidos se emplearon como referencias en un ensayo proteómico para muestras de proteína de *P. lunula*. La búsqueda de los espectros obtenidos contra las bases de datos Trinity_Q20

L100_k25_NC20 y Trinity_Q20 L100_k31_NC20 arrojó el mayor número de proteínas identificadas. Estos resultados, junto con los análisis de representatividad y estructura comentados anteriormente, nos permitieron elegir el ensamblaje Trinity_Q20 L100_k25_NC20 (414,295 contigs (185,177> 400 bases; 219,203,407 pb; 55.38% GC) como el más adecuado para los perfiles de identificación obtenidos por análisis MS [Material Suplementario 2].

3.2 ANÁLISIS PROTEÓMICO

Para inferir un proteoma representativo optimizado de las secuencias de transcripción ensambladas en Trinity_Q20 L100_k25_NC20, se realizó un BLASTx contra la base de datos "nr", que representa todas las secuencias de proteínas públicas, no redundantes. Como resultado, se encontraron coincidencias para alrededor de 100,000 secuencias (umbral: valor-e <e-15). Esta información fue muy útil, ya que se realizó una anotación funcional adicional de estas transcripciones coincidentes, así como para identificar su marco de lectura, permitiendo la identificación de un conjunto de secuencias de codificación de *P. lunula*.

Para las 300,000 secuencias restantes que no mostraron una homología clara, se identificaron todos los ORF posibles de los 6 marcos de lectura, y se retuvo la secuencia de proteínas codificada por el ORF más largo, junto con el conjunto de secuencias de codificación previamente identificadas, para luego buscar dentro del ensayo proteómico. Aunque algunas de estas secuencias de proteínas recientemente inferidas podrían ser artefactos (consecuencia de traducciones de ARN no codificantes o artefactos de ensamblaje / secuenciación) que introducen ruido en la base de datos proteómica, consideramos de gran relevancia el análisis de estas nuevas proteínas putativas que podrían estar presentes y que aún no se han reportado en bases de datos públicas. Esto podría ser de especial importancia en especies claramente subrepresentadas en las bases de datos, como ocurre con la falta de estudios moleculares para todo el orden Pyrocystales, así como para muchos dinoflagelados en general.

Todas las secuencias generadas en ambos grupos, con o sin homología contra "nr", si eran mayores de 50 aa, se conservaron y se creó el proteoma de referencia. El software Proteome Discoverer 2.1 (Thermo Fisher Scientific), con el algoritmo Sequest HT, se utilizó para buscar todos los espectros que se obtuvieron en la muestra de proteína en el análisis MS / MS (17.461 péptidos) contra ese proteoma predicho, y con un umbral de FDR <0.01, se obtuvieron 3.182 identificaciones. Se estudió la distribución bidimensional de las proteínas identificadas a través del ensayo proteómico dentro de las proteínas totales presentes en el proteoma predicho, de acuerdo con su peso molecular (MW) y punto isoelectrico (Ip) (Fig. 8). Los datos de espectrometría de masas se depositaron en el Consorcio ProteomeXchange a través del repositorio asociado PRIDE (Vizcaino *et al.*, 2016), con el identificador del conjunto de datos PXD011555.

3.3 ANOTACIÓN FUNCIONAL

Usando el software BLAST2GO, todas las proteínas identificadas se anotaron funcionalmente de acuerdo con las categorías GO, los códigos de enzimas y los dominios InterPro. Por lo tanto, las Figuras 9 y 10 muestran las funciones principales identificadas dentro de las proteínas identificadas de acuerdo con el proceso biológico GO (BP) y las categorías de función molecular (MF), respectivamente. Una lista completa de proteínas identificadas, junto con toda la información funcional asociada se informa en el Material Suplementario 3. Con respecto a la categoría de proceso biológico (BP), el primer lugar está ocupado por proteínas relacionadas con procesos de óxido-reducción, seguidas por proteínas relacionadas con el proceso de traducción y biogénesis de los ribosomas. A nivel de la función molecular (MF), se encontró que las moléculas de unión a ATP eran más abundantes entre el proteoma, seguidas por los constituyentes estructurales de unión a ribosomas y a iones metálicos.

Todos los códigos enzimáticos asociados se ubicaron en la base de datos KEGG, utilizando la herramienta KEGG-mapper, para identificar las vías en las que las proteínas identificadas podrían estar actuando [Material Suplementario 4]. Atendiendo a las especies que contribuyen a las asignaciones funcionales, *S. microadriaticum* fue la especie de mayor éxito, seguida de *Vitrella brassicaformis*, *Perkinsus marinus* y *Heterocapsa triquetra* (Fig. 11). Aunque este análisis puede sufrir un sesgo por la abundancia de secuencias de cada especie presente en las bases de datos públicas, nuestros resultados parecen consistentes con la relación evolutiva entre Dinoflagellata, Chromerida y Perkinzoa (alveolatos) (Okamoto y Hasting, 2003; Lin, 2011; Janouskovec *et al.*, 2017).

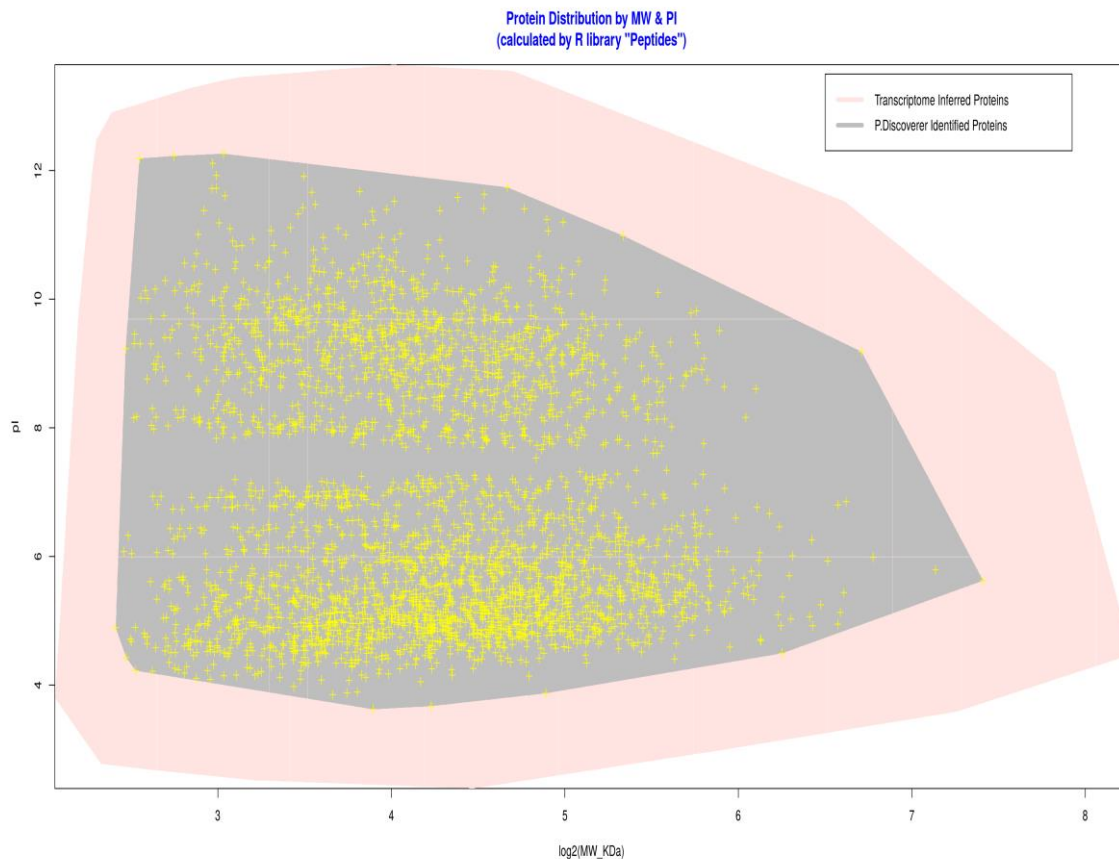


Figura 8. Distribución de las proteínas predichas e identificadas de acuerdo con su peso molecular (MW) y el punto isoeléctrico (pI) calculado por los péptidos de la biblioteca R.

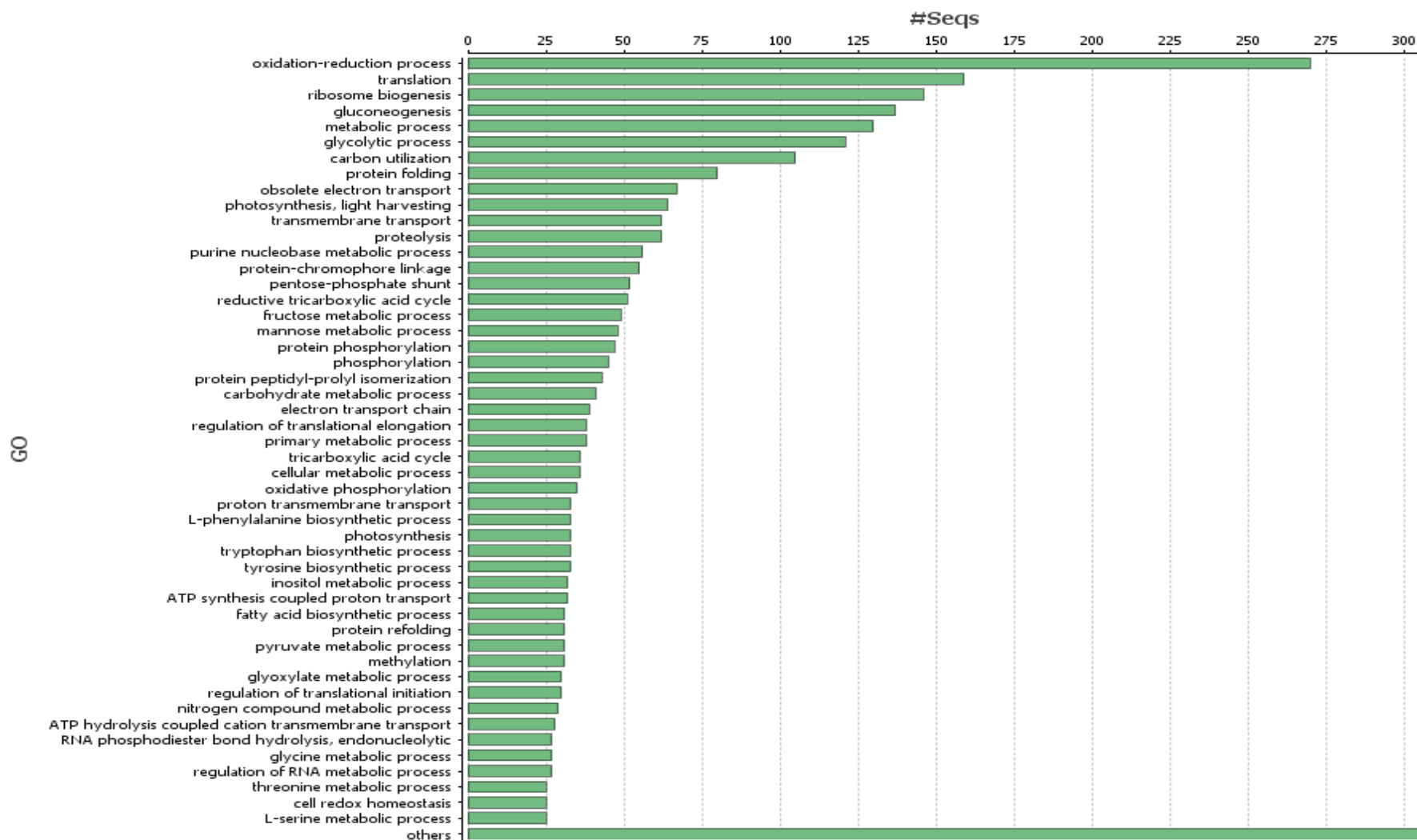


Figura 9. Análisis funcional de GO de las secuencias de *P. lunula* según la clasificación del proceso biológico (BP).

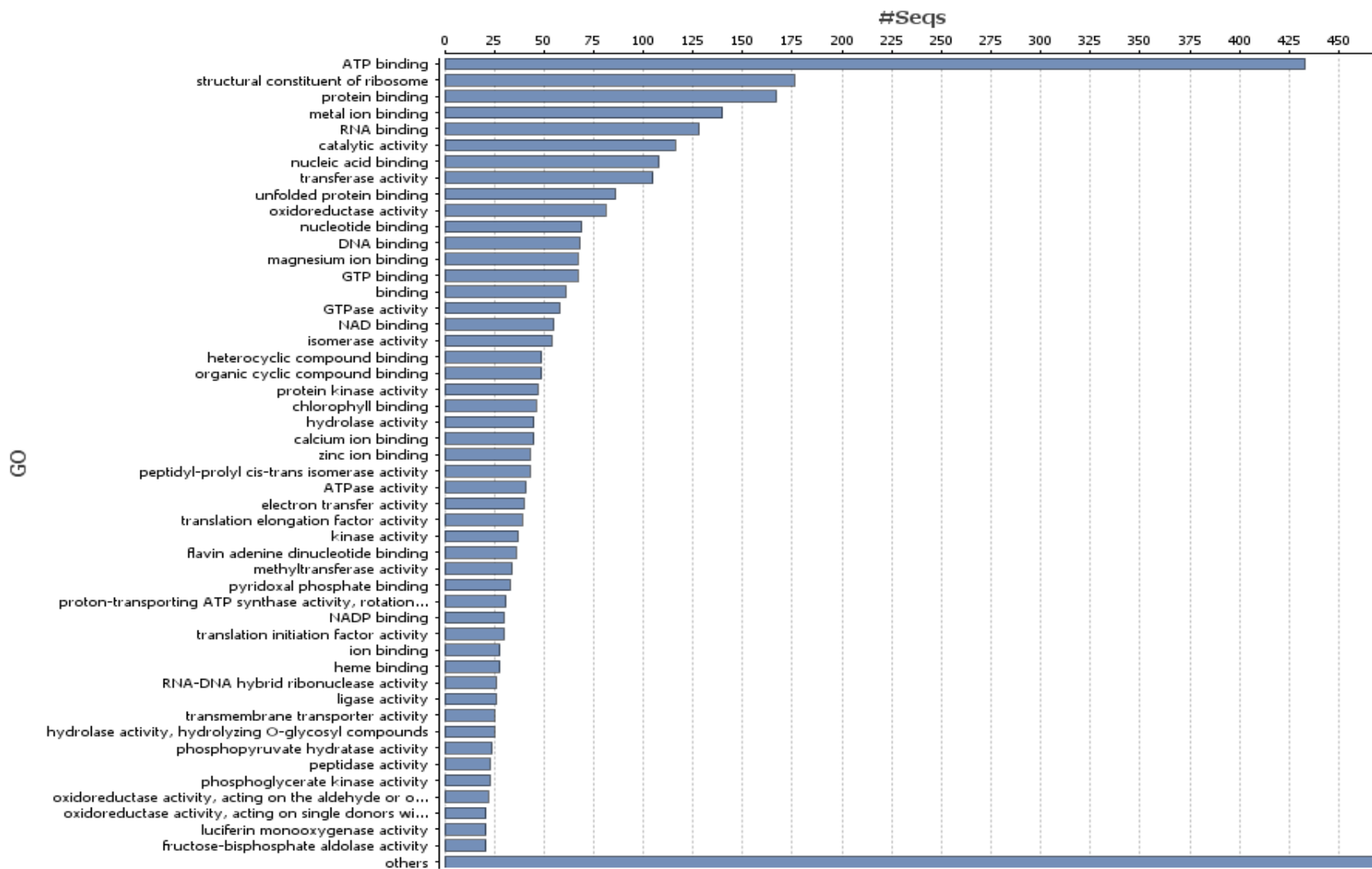


Figura 10. Análisis funcional de GO de las secuencias de *P. lunula* según la clasificación de la función molecular (MF).

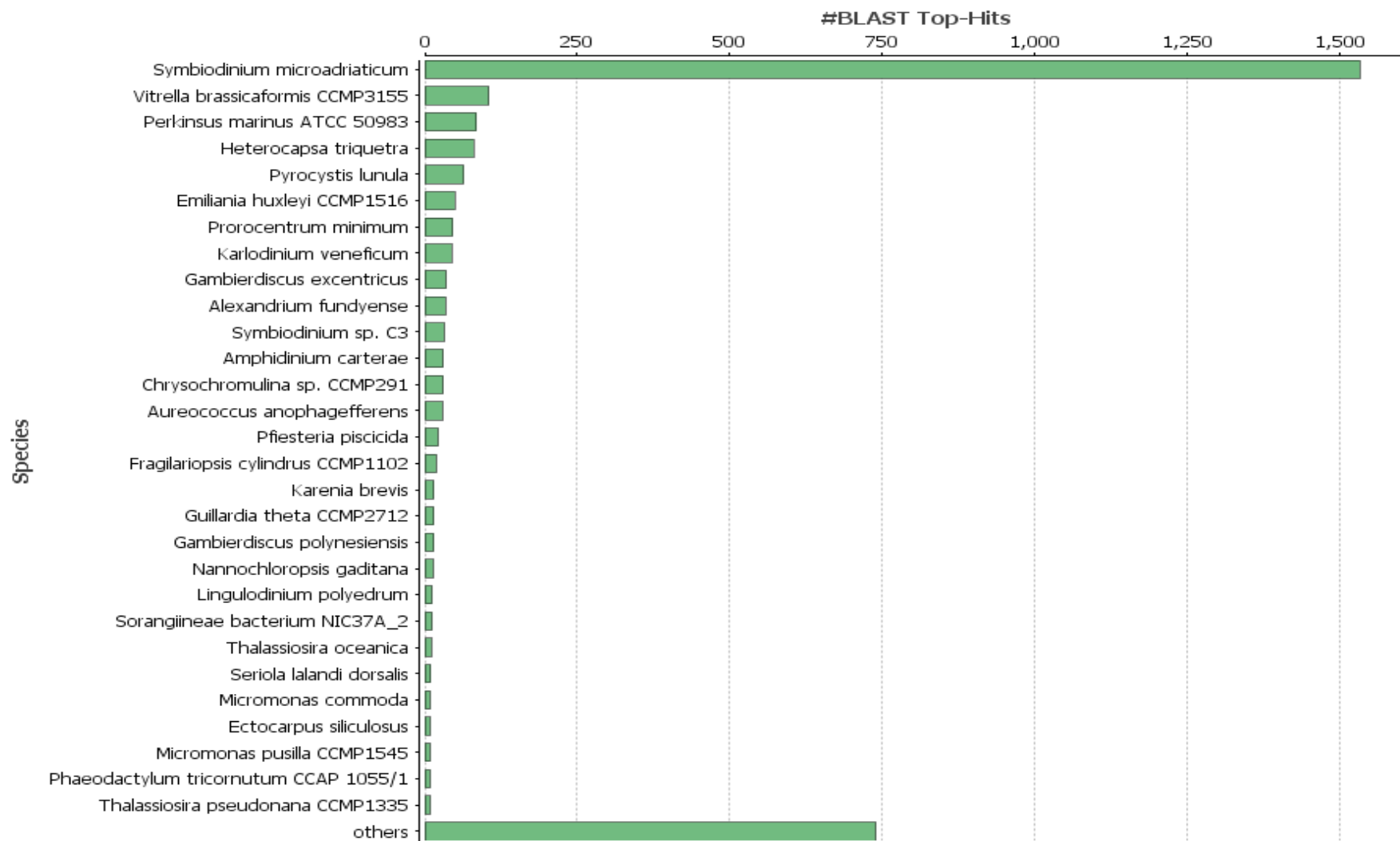


Figura 11. Distribución de proteínas por especies para las secuencias de *P. lunula* según BLAST2GO.

4. DISCUSIÓN

4.1 ANÁLISIS ESTRUCTURALES Y FILOGENÉTICOS

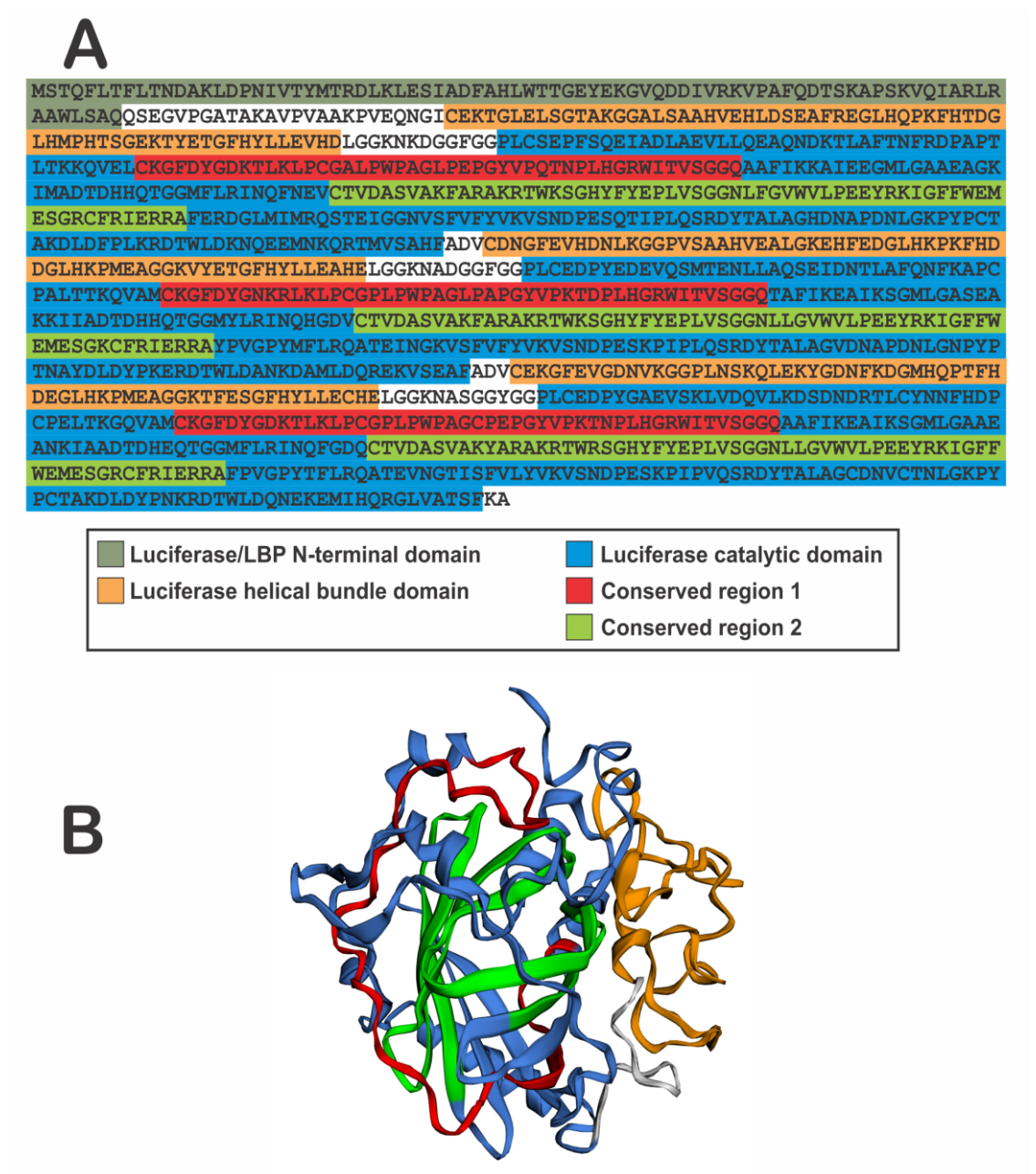
4.1.1 Luciferasa (LCF)

En *P. lunula*, como en todos los Gonyaulacales estudiados hasta ahora, LCF (MW = 137 kDa) tiene un solo dominio terminal LCF / LBP N y 3 dominios catalíticos, precedidos por dominios de haz de helicasa (Fig. 12), con cada dominio catalítico, de aproximadamente 46 kDa, siendo enzimáticamente activo, y su secuencia de codificación está dispuesta, dentro del genoma, en varias copias en tándem (Morishita *et al.*, 2002; Liu *et al.*, 2004). En *P. lunula* hay tres isoformas diferentes del gen LCF: LCFa (GenBank AF394059.1), LCFb (GenBank AF394060.1) y LCFc (GenBank AF394061.1) (Okamoto *et al.*, 2001), y la estructura primaria está muy conservada (Valiadi *et al.*, 2012).

El alineamiento de las secuencias de los tres dominios catalíticos en las especies disponibles de Gonyaulacales muestra la presencia de 2 dominios altamente conservados (Fig. 13) que deberían estar asociados a la función catalítica de la enzima. El logotipo de secuencia muestra 87.2 y 93.8 de identidad en las secuencias de aminoácidos en las regiones 1 y 2, respectivamente. El análisis filogenético de los dominios LCF de las especies de Gonyaulacales, usando LCF de *N. scintillans* como un grupo externo, que solo contiene un dominio catalítico, muestra un patrón evolutivo muy similar para cada uno de los tres dominios (Fig. 14).

La relación filogenética entre las especies se conserva, pero tiene algunas diferencias con un estudio publicado anteriormente (Fig. 15) (Orr *et al.*, 2012), ya que nuestro árbol muestra una relación más estrecha entre las especies del género *Alexandrium*, *P. reticulatum* y *C. horrida* con el género *Pyrocystis*, que es lo opuesto a lo que se informa en la relación filogenética construida usando las secuencias de 8 genes diferentes (28S, 5.8S, 18S, *cox1*, *cob*, *beta-tubulina*, *actina*, y *hsp90*). Es probable que la formación de los tres dominios ocurriera por duplicación del dominio 1 y antes de la diversificación de los Gonyaulacales, lo que es una característica común en todas las especies secuenciadas, pero no hay evidencia de si los tres dominios están presentes

fuera del grupo de Gonyaulacales, ya que no hay ningún informe de la secuencia completa de LCF en ningún otro dinoflagelado fuera de este orden, excepto *N. scintillans*.



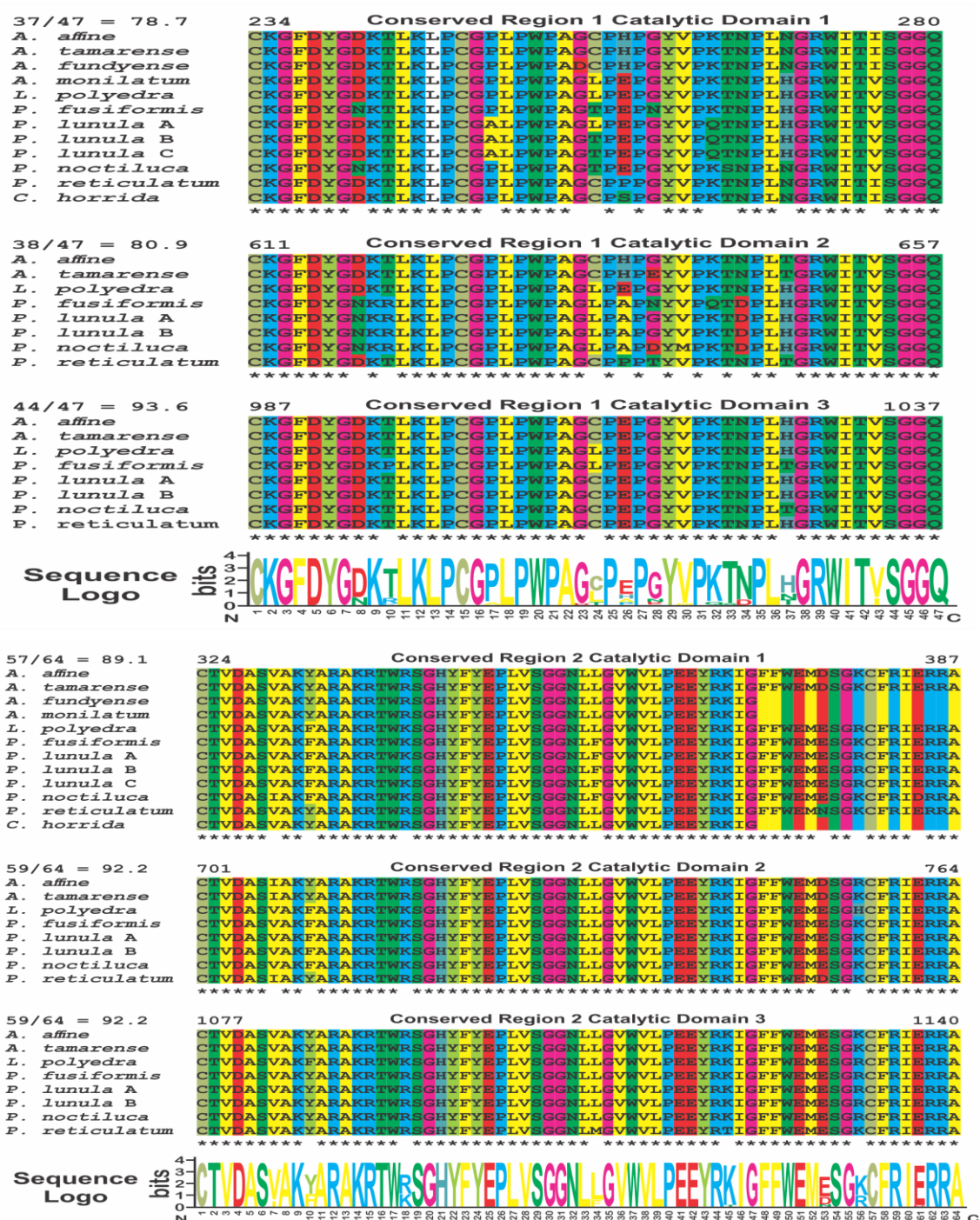


Figura 13. La alineación de la secuencia de las regiones conservadas 1 (A) y 2 (B) encontradas en la enzima LCF se compara entre los tres dominios catalíticos y las secuencias disponibles de las especies de *Alexandrium*, *Lingulodinium*, *Pyrocystis* y *Protoceratium*. Los números en la parte superior de cada región corresponden a la posición de aminoácidos en la proteína de *P. lunula* (GenBank AAL40676). Los asteriscos muestran los aminoácidos conservados en cada posición. La coloración de los aminoácidos se realizó de acuerdo con el mismo patrón mostrado en el software MEGA7. El logotipo de secuencia se realizó utilizando WebLogo, como se describe en el texto.

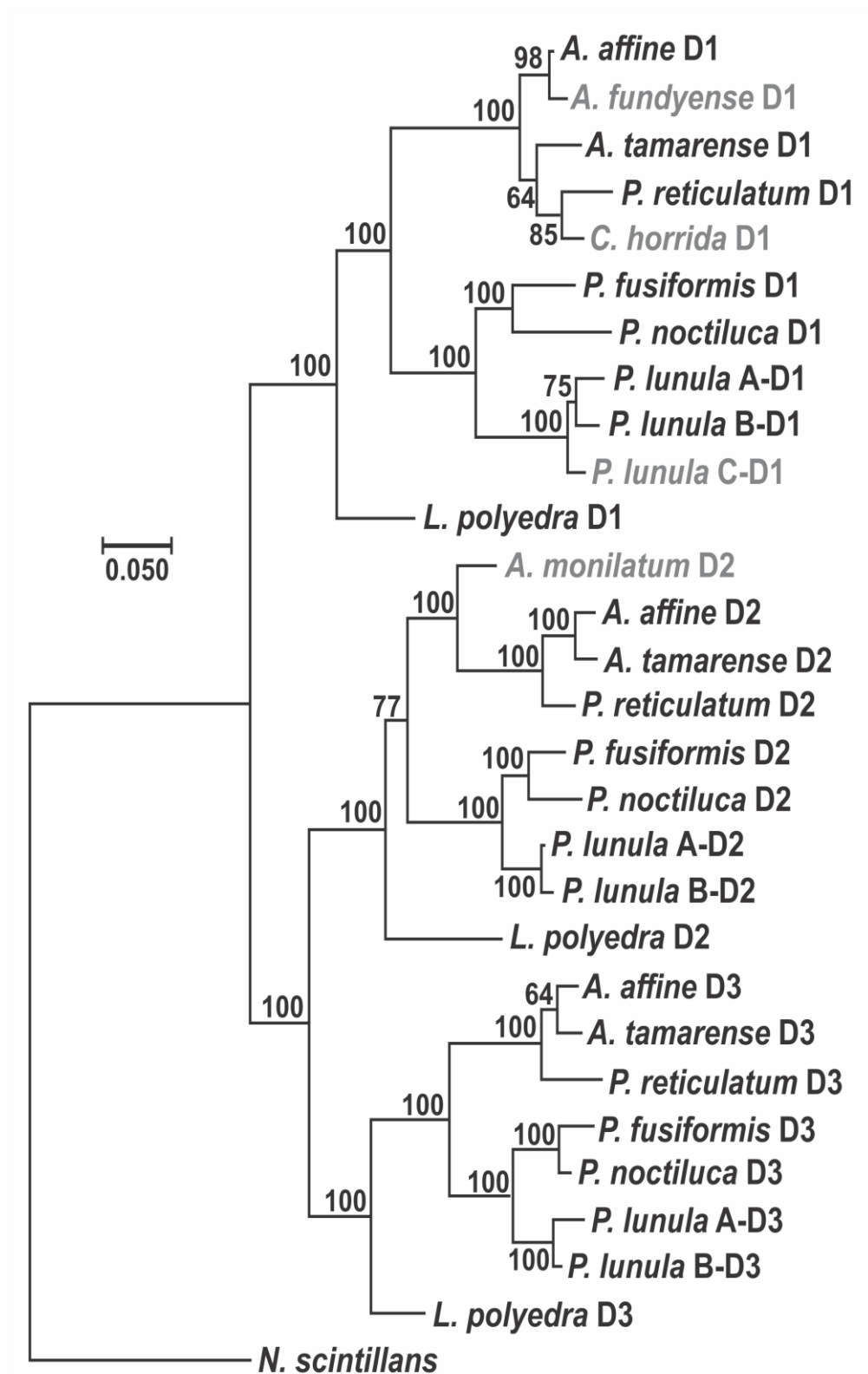


Figura 14. Análisis filogenético molecular por un método bayesiano, que muestra el patrón evolutivo de los tres dominios catalíticos de LCF en las especies de dinoflagelados con secuencias de proteínas disponibles, utilizando el único dominio catalítico de *N. scintillans* como grupo externo.

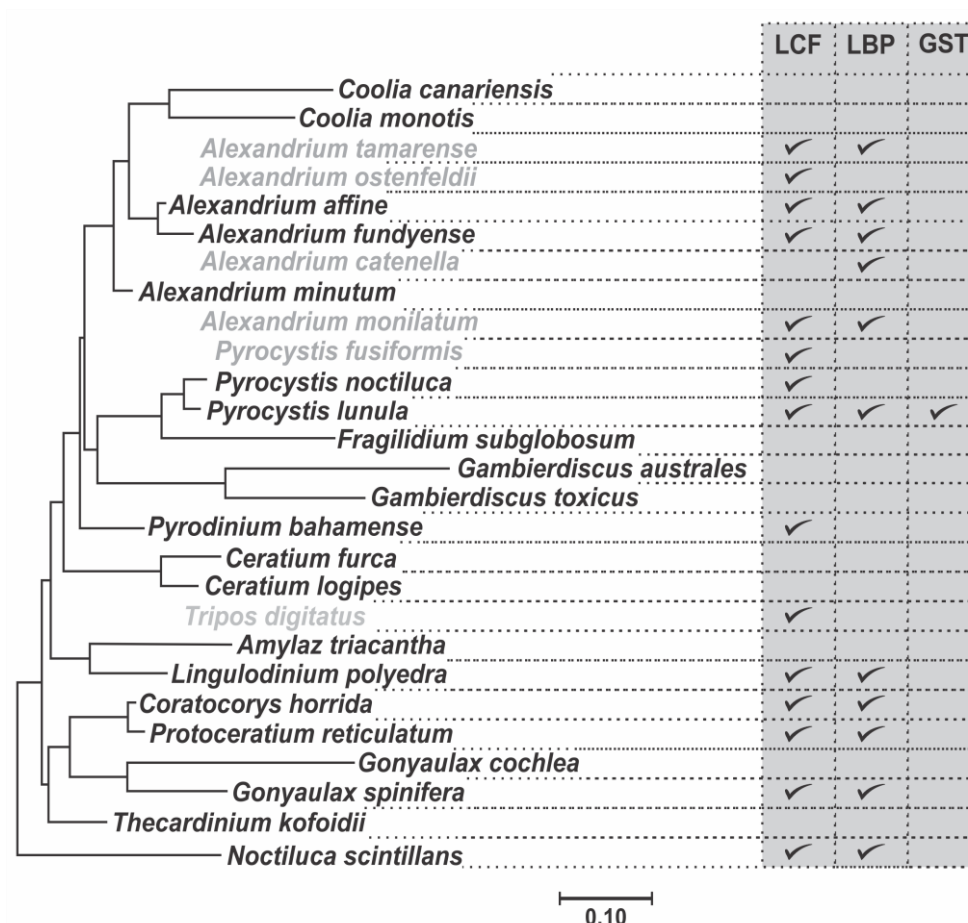


Figura 15. Árbol filogenético de los dinoflagelados Gonyaulacales inferidos de los tres genes de ADNr, dos genes mitocondriales y cuatro proteínas nucleares, modificados de Orr *et al.* (2012). El árbol se reconstruyó con inferencia bayesiana (MrBayes) e incluyó a los diferentes grupos dentro de los dinoflagelados, pero aquí solo se muestran los Gonyaulacales. Las especies en gris no se incluyeron en el análisis original de Orr *et al.* (2012). La tabla muestra en cuáles de estas especies se han reportado las secuencias de las proteínas LCF, LBP y GST.

En *L. polyedra*, se informó que las actividades eran máximas a pH 6.3, disminuyendo casi a cero a pH 8. Cuatro histidinas conservadas intramolecularmente en el dominio del haz helicoidal LCF están relacionadas con la pérdida de actividad a pH alto (Li *et al.*, 2001). Estos residuos de histidina también se conservan en *P. lunula* (Li *et al.*, 1997; Okamoto *et al.*, 2001; Morishita *et al.*, 2002; Liu *et al.*, 2004; Valiadi *et al.*, 2012), así como en las otras especies analizadas aquí, al igual que todos los dominios de paquetes helicoidales (Fig. 16). Las dos primeras de las histidinas conservadas están precedidas por glicina y leucina y las dos últimas se encuentran en una región de 10 aminoácidos que muestran una conservación muy alta.

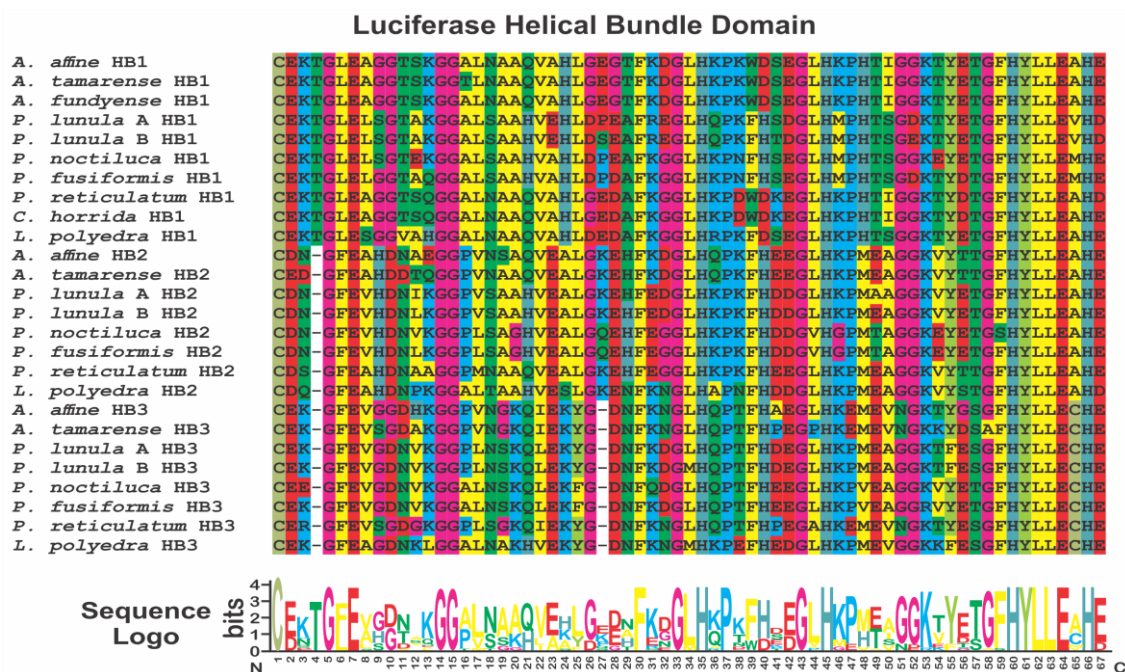


Figura 16. Alineación de secuencia de los tres dominios del paquete helicoidal contenidos en la enzima LCF de las especies de *Alexandrium*, *Lingulodinium*, *Ceratocorys*, *Pyrocystis* y *Protoceratium* que tienen su proteína publicada en GenBank. La coloración de los aminoácidos se realizó de acuerdo con el mismo patrón mostrado en el software MEGA7. El logotipo de secuencia se realizó utilizando WebLogo, como se describe en el texto.

La estructura 3D de los aminoácidos que codifican una enzima determinan cómo se une el sustrato y qué reacción química puede ocurrir. Se ha reportado la estructura del tercer dominio de LCF de *L. polyedra* (Schultz *et al.*, 2005), y de forma general se puede dividir en dos estructuras principales: 1) el barril, que comprime el núcleo de la enzima y tiene el sitio activo donde se produce la oxidación de la luciferina y la producción de luz, y 2) la tapa del barril, que está cerrada por un paquete regulatorio de tres hélices en la parte superior de la estructura. Esta estructura es muy similar a la reportada aquí para *P. lunula* (Fig. 12), con las regiones conservadas 1 y 2 ubicadas en las paredes del barril. Los estudios de dinámica molecular indican que cuando el pH cae cerca de 6, las histidinas reguladoras N-terminales comienzan un cambio conformacional en el paquete de tres hélices (Wilson y Hasting, 2013). La tapa del barril se abre y permite que la luciferina acceda al sitio activo, donde se oxida. En esta estructura, las moléculas de agua ocupan el sitio activo que tomaría la luciferina durante la reacción bioluminiscente. Los residuos de histidina N-terminal que controlan la actividad mediante el pH están

en la interfaz de las hélices en el paquete. El pKa de la histidina generalmente oscila entre 6 y 6.5, lo que indica que el estado de protonación de estos residuos cambia cuando el pH cae por debajo de 6.5, cuando la LCF está activa. Para LCF, el estado de protonación de los residuos de histidina ubicados fuera del sitio activo regula la actividad (Wilson y Hasting, 2013).

Recientemente, Donnan *et al.* (2017) estudiaron los cambios estructurales relacionados con la activación de LCF tras la acidificación. La protonación de algunos residuos, incluidas las histidinas conservadas intramolecularmente y la díada H1064 / H1065 (dentro del dominio catalítico), se correlacionan con un cambio estructural a gran escala en el que los dominios del haz helicoidal se reagrupan para permitir el acceso de la luciferina al sitio activo. Paralelamente, el barril β se expande y una supuesta base del sitio activo, E1105, se mueve a su posición comenzando la reacción de catálisis.

Análisis mediante PCR con ADN genómico de *P. lunula* mostraron que, entre las isoformas de LCFa, LCFb y LCFc, las dos primeras están en repetición en tándem, pero no hay un promotor identificable en los espaciadores intergénicos, lo que ha llevado a sugerir que las repeticiones de genes en tándem pueden constituir un mecanismo de transcripción policistronica, similar a las estructuras genéticas encontradas en *Trypanosoma* (Jackson, 2007). Esto podría aumentar la eficiencia de la transcripción / traducción de LCF.

En algunos casos, como en *L. polyedra*, la abundancia de proteína LCF muestra un ritmo circadiano, sintetizándose en la fase de oscuridad temprana y destruyéndose en el transcurso del período de luz. Sin embargo, existen claras diferencias en el sistema de *P. lunula*, donde el parpadeo, pero no el brillo tiene un ritmo circadiano. Además, las cantidades de luciferina y LCF son constantes durante el día y la noche en *P. lunula* (Colepicolo *et al.*, 1993; Knaust *et al.*, 1998). En relación a por qué existe en algunas especies, como *L. polyedra*, un mecanismo de síntesis *de novo* de los componentes moleculares del sistema bioluminiscente, lo que resulta contraintuitivo desde el punto de vista del gasto energético involucrado, Hasting (2013) señala que puede representar un

mecanismo para conservar el nitrógeno. De esta forma los aminoácidos liberados por la hidrólisis de una proteína, como LCF y LBP, podrían estar disponibles para la biosíntesis de diferentes proteínas a lo largo del ciclo circadiano.

En el caso de *P. lunula*, que también exhibe un ritmo circadiano de bioluminiscencia y no destruye ni vuelve a sintetizar LCF, puede estar relacionado con su ecología, ya que esta especie exhibe una migración vertical diaria a aguas más profundas y oscuras donde el nitrógeno puede estar fácilmente disponible. Se presume también que el alto número de copias de LCF resultó de la necesidad de una gran cantidad de esta proteína, que ejerce una presión selectiva para la retención de copias genéticas duplicadas (Li *et al.*, 1997, Lin, 2011). Otra posible razón para la presencia de tres sitios catalíticos activos en una sola molécula es que permite que la actividad sea mayor sin un aumento en la presión osmótica coloidal dentro del scintillon (Wilson y Hasting, 2013).

4.2 NUEVAS PROTEÍNAS Y MINERÍA DE DATOS PROTEÓMICOS

Las microalgas han demostrado su gran potencial biotecnológico, no solo desde el punto de vista de su uso como biorefinerías, sino también por la cantidad de nuevos genes que han sido detectados en estos organismos. De las 3.182 proteínas detectadas en *P. lunula*, 175 (5,49 %) resultan ser secuencias que no tienen homología con ninguna proteína identificada hasta ahora en las diferentes bases de datos [Material Suplementario 5]. Dentro de este grupo, 59 se destacan ($PSM \geq 5$) y deben considerarse con un alto nivel de confianza como proteínas novedosas. Esto no es inesperado ya que los conjuntos de datos de transcriptómica de dinoflagelados han demostrado, entre otras cosas, que en algunos casos más del 50 % de los genes no concuerdan con ningún gen documentado en ningún organismo (Lin, 2011). Muchos genes parecen únicos a los dinoflagelados. Por ejemplo, aunque muchos de los genes regulados por mecanismos redox identificados en *P. lunula* se conservan, la mayoría (~ 68 %), no comparten similitudes significativas con otros genes en la base de datos (Okamoto y Hasting, 2003).

Siguiendo este paradigma nos propusimos explorar el proteoma de *P. lunula* en busca de proteínas que tuviesen valor desde una perspectiva biotecnológica o cuya detección ayudara a describir mejor la biología básica de esta especie. *P. lunula* ha constituido un organismo modelo para el estudio de la bioluminiscencia y es precisamente en relación con este mecanismo que hemos logrado identificar algunos componentes que se pensaban ausentes en el sistema bioluminiscente de este género de dinoflagelados en particular. Este es el caso de la proteína LBP, lo cual constituye una nueva y más acertada perspectiva con relación al mecanismo molecular responsable del evento bioluminiscente en estos organismos.

4.2.1 Proteína de unión a la luciferina (LBP)

La presencia del gen LBP es bien conocida en *L. polyedrum* (Fogel y Hasting, 1971; Schmitter *et al.*, 1976) y *N. scintillans* (Liu y Hastings, 2007), así como en los géneros *Gonyaulax*, *Ceratocorys*, *Protoceratium*, y *Alexandrium* (Hackett *et al.*, 2005; Erdner y Anderson, 2006; Uribe *et al.*, 2008; Jaeckisch *et al.*, 2011); por lo tanto, este gen parece ser un componente integral de la maquinaria molecular bioluminiscente estándar en los dinoflagelados. Los análisis de las secuencias codificantes de LBP mostraron una alta variación intra-cepa entre las copias genéticas, revelando una familia de genes muy diversa que comprende múltiples tipos de genes en algunos organismos (Valiadi e Iglesias, 2014). De hecho, se ha informado que *Pyrocystis* carece de la expresión de LBP, cuyos extractos de proteínas se analizaron con un anticuerpo universal para LBP (Schmitter *et al.*, 1976; Knaust *et al.*, 1998); sin embargo, esto no podría tomarse como una prueba concluyente de la ausencia de la proteína LBP, debido al alto grado de variabilidad informado para este gen. Actualmente se desconoce la cantidad de especies de dinoflagelados bioluminiscentes que utilizan LBP (Valiadi y Iglesias, 2013). Sin embargo, información emergente muestra evidencia sustancial de que la LBP, con su compleja historia evolutiva, es una parte importante de los sistemas de bioluminiscencia en dinoflagelados (Valiadi e Iglesias, 2014).

Se ha demostrado que cebadores diseñados para amplificar LBP (DinoLBP) de dinoflagelados produjeron productos de PCR de once cepas de las dieciocho probadas, revelando que la LBP estaba presente en todos los miembros de Gonyaulacales con la excepción de los géneros *Ceratium* y *Fragilidium*. Es importante tener en cuenta que, debido a las grandes diferencias de la LBP entre especies, la eficiente amplificación de LBP de *Alexandrium monilatum*, *Ceratocorys horrida* y *Gonyaulax spinifera* solo fue posible después de una reducción en la rigurosidad de la PCR. Además, los cebadores DinoLBP amplificaron la LBP en todas las especies que se sabe que la contienen, excepto en *N. scintillans*, que contiene una secuencia muy divergente de tipo LBP (Liu y Hastings, 2007).

Se ha reportado la distribución y diversidad de LBP en varios dinoflagelados gonyaulacoides estrechamente relacionados, adicionalmente a los primeros informes de LBP en *C. horrida*, *G. spinifera* y *P. reticulatum*, así como las primeras secuencias parciales de LBP aisladas de *Alexandrium*, *A. fundyense*, *A. monilatum* y *A. tamarense* (Valiadi e Iglesias, 2014). Estos resultados también concuerdan con la ausencia de LBP en *Pyrocystis spp.*; sin embargo, la falta de detección de LBP en *N. scintillans*, que se sabe que contiene esta secuencia, sugiere que el resultado negativo de la PCR en *Ceratium longipes*, *Fragilidium cf. subglobosum*, *Protoperidinium crassipes*, e incluso en *P. lunula*, no son concluyentes para la ausencia de este gen. A pesar de muchos esfuerzos, las grandes diferencias en las secuencias de LBP entre especies han impedido la síntesis de cebadores universales para la detección de la LBP, tal y como se hizo anteriormente en el caso de LCF (Valiadi *et al.*, 2012).

Las secuencias de LBP han demostrado ser una familia de genes muy grande y muy diversa (Machabee *et al.*, 1994; Tanikawa *et al.*, 2004). Además, los análisis filogenéticos de *Alexandrium spp.* también sugieren la presencia de secuencias de más de un tipo de LBP, similar a las observaciones previas con sus secuencias LCF (Valiadi *et al.*, 2012). Múltiples tipos de isoformas en genes involucrados en la bioluminiscencia parecen ser comunes tanto en la LBP como en la LCF para varias especies de Gonyaulacales (Machabee *et al.*, 1994;

Okamoto *et al.*, 2001; Valiadi *et al.*, 2012), creando una mayor divergencia entre las copias del gen de la LBP en estos genomas.

Estudios en *N. scintillans* también revelaron que la LBP se presenta en diversas formas como resultado de eventos evolutivos significativos, como la fisión o fusión de genes. El caso de *N. scintillans* es inusual ya que LCF y LBP se encuentran como dos dominios en un solo gen, mientras que normalmente se estructuran como genes separados en todas las especies fotosintéticas estudiadas hasta esa fecha (Valiadi e Iglesias, 2014). Se ha sugerido que el origen de la estructura híbrida de esta LCF/LBP es debida a la fusión de los dos genes de especies fotosintéticas en *N. scintillans* o, por el contrario, la fisión del gen de *N. scintillans* en la especie fotosintética. El gen combinado LCF/LBP (2.396 pb) descrito previamente para este organismo (Liu y Hastings 2007; Valiadi e Iglesias, 2014), consisten en parte de la región N-terminal y el dominio LCF seguidos por el dominio LBP.

Un análisis detallado de nuestros resultados ha revelado una situación similar, pudiendose detectar dos isoformas diferentes de LBP, una versión que parece corresponder al gen individual LBP (GenBank MN259726) [Material Suplementario 6], muy similar (longitud 422 nt, cobertura 95 %, identidad 75 %) a la informado para *A. tamarense* (GenBank AFN27006.1) (Valiadi e Iglesias, 2014), y otra que representaría la versión híbrida entre LCF / LBP (GenBank MN259727) [Material Suplementario 7], más similar a las secuencias reportadas para *L. polyedra* (GenBank AAA29164.1 y AAA29163.1) (longitud 1.116 nt, cobertura 96 %, identidad 55 %) (Machabee *et al.*, 1994) y *A. cantenella* (GenBank: ABY78836.1) (longitud 1116 nt, cobertura 95 %, identidad 54 %) (Uribe *et al.*, 2008). Estos resultados se corroboraron mediante PCR, secuenciación (Sanger) y análisis BLASTx (Altschul *et al.*, 1997).

De hecho, una situación similar ocurre en el caso del gen LCF en *P. lunula*, donde se han descrito hasta tres isoformas: LCFa (GenBank AF394059.1), LCFb (GenBank AF394060.1) y LCFc (GenBank AF394061. 1) (Okamoto *et al.*, 2001). El elemento característico de la versión híbrida LCF/LBP es la presencia del dominio N-terminal LCF/LBP (pfam05295,

Luciferasa_N); sin embargo, este dominio no es la parte catalítica de la proteína y se ha sugerido que puede mediar una interacción entre LBP y LCF o su asociación con la membrana vacuolar (Valiadi e Iglesias, 2014).

Además, en *N. scintillans*, ambos fragmentos contienen una región génica N-terminal idéntica con una longitud de 108 pb. Una alineación de las dos secuencias de ADNc traducidas reveló que, en el fragmento más corto, la región N-terminal, condujo directamente al dominio LBP en ausencia de un dominio LCF. Por lo tanto, estas secuencias, que compartían la región N-terminal y el dominio LBP, pero no el dominio LCF, corresponden a dos genes diferentes: LCF/LBP híbrido y LBP individual. Mediante el uso de ADNc, también se confirmó que *N. scintillans* expresa ambos genes (Valiadi e Iglesias, 2014).

Se ha especulado que una segunda LBP en *N. scintillans*, podría tener un papel en la unión de luciferina en los escintillones o podría actuar para almacenar luciferina en el citoplasma (Liu y Hastings, 2007). Si la LCF/LBP de *N. scintillans* es ancestral a la LCF y LBP de las especies fotosintéticas, como lo sugiere su filogenia general (Orr *et al.*, 2012), la LBP individual de *N. scintillans* podría haberse formado por un empalme de ARNm de LCF / LBP y posterior retro-transposición en el genoma, lo cual se ha reportado anteriormente en dinoflagelados (Bachvaroff y Place, 2008; Slamovits y Keeling, 2008). La región génica N-terminal idéntica en ambos genes apoya esta hipótesis.

Tener genes separados para LCF y LBP podría permitir la regulación diferencial de cada gen. Esto puede ser ventajoso si la LBP necesita estar estequiométricamente en proporción a la cantidad de luciferina (Fogel y Hastings 1971), mientras que la LCF, que tiene una capacidad catalítica triple en especies fotosintéticas, puede reutilizarse (Valiadi e Iglesias, 2014). El dominio terminal LCF N muestra una alta conservación de los aminoácidos entre las proteínas LCF y LBP y entre las especies fotosintéticas con secuencias disponibles (Fig. 17), destacando el importante papel que este dominio tiene para la función de ambas proteínas en el sistema de

bioluminiscencia. Es interesante la alta frecuencia de aminoácidos polares cargados y no cargados, lo que podría sugerir un papel en la interacción de estas proteínas con la superficie polar de los fosfolípidos en las membranas.

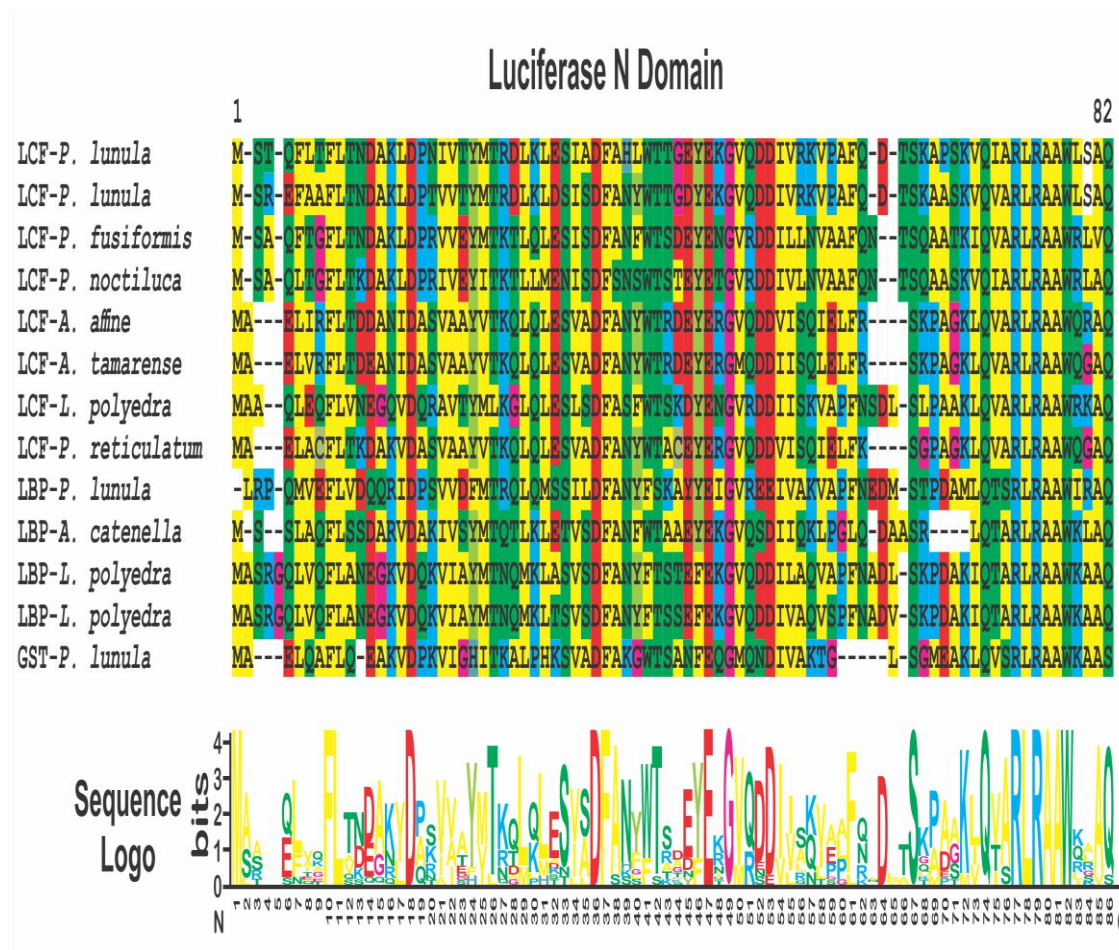


Figura 17. Alineación de secuencia del dominio terminal LCF / LBP N contenido en la enzima LCF, LBP y GST de las especies de *Alexandrium*, *Lingulodinium*, *Pyrocystis* y *Protoceratium* que tienen su proteína publicada en GenBank. La coloración de los aminoácidos se realizó de acuerdo con el mismo patrón mostrado en el software MEGA7. El logotipo de secuencia se realizó utilizando WebLogo, como se describe en el texto.

El análisis filogenético de las secuencias de LBP disponibles en GenBank muestra un patrón evolutivo muy similar al obtenido en el análisis de los dominios catalíticos de LCF (Fig. 18), que difiere del patrón visto en las Gonyaulacales al analizar 8 genes (Fig. 15) (Orr *et al.*, 2012). Este análisis muestra que las secuencias LBP de *C. horrida* y *P. reticulatum* están muy relacionadas con las especies de *Alexandrium*, mientras que la secuencia de *P. lunula* parece ser la más divergente.

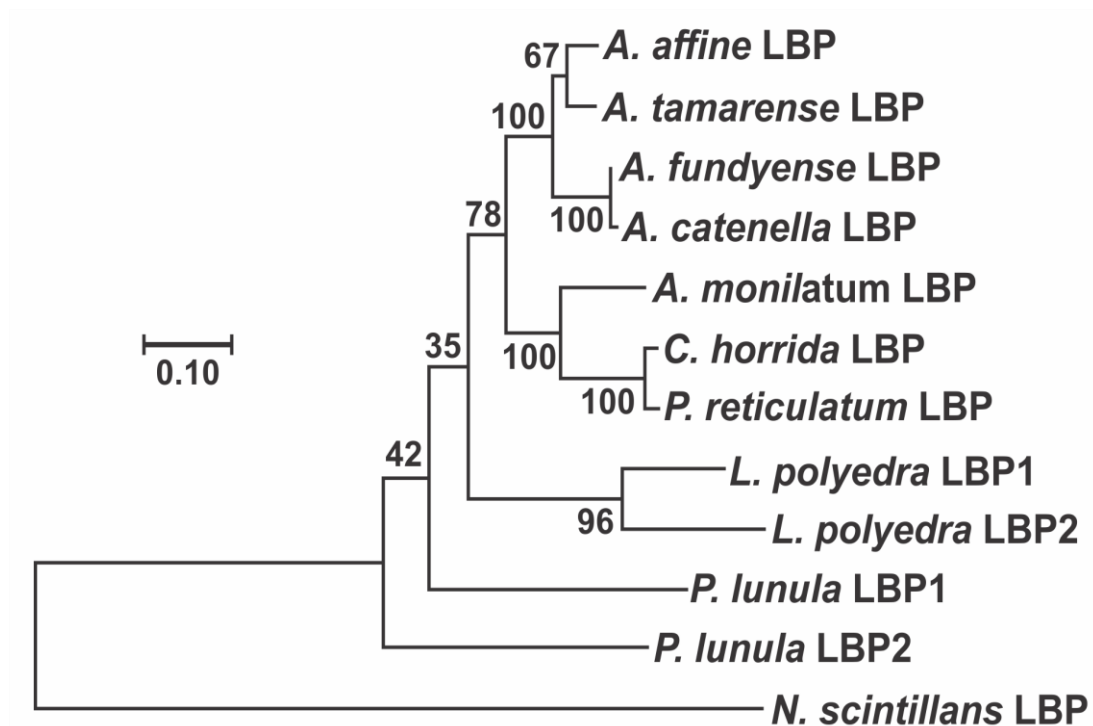


Figura 18. Análisis filogenético molecular por un método bayesiano, que muestra el patrón evolutivo de la proteína LBP de las especies de *Alexandrium*, *Lingulodinium*, *Pyrocystis* y *Protoceratium* que tienen su proteína publicada en GenBank, utilizando la secuencia de *N. scintillans* como grupo externo..

4.2.2 Glutación-S-transferasa (GST) y la escurridiza luciferina

Otro hallazgo interesante lo constituye la detección de una isoforma de la proteína GST que contiene el dominio Luciferase_N pfam05295. Esta proteína podría estar relacionada con el proceso de reciclaje de la luciferina al actuar como una reductasa que invertiría el estado de oxidación de la oxiluciferina, reciclándola así hasta luciferina. En este proceso estaría involucrado un mecanismo auxiliar a la cadena de electrones no descrito hasta la fecha. Esto a primera vista podría constituir una hipótesis inverosímil; sin embargo, si también tomamos en cuenta la presencia del mecanismo de la bomba de rodopsina, detectado también en nuestros resultados, y cuyo descubrimiento puso por primera vez a prueba el concepto de que es posible obtener ATP a partir de energía lumínica a través de una vía metabólica distinta a la fotosíntesis; esto nos demuestra el grado de plasticidad de las estrategias adaptativas desplegadas por estos organismos y hasta que punto el estudio de los mismos puede cambiar nuestra concepción de los paradigmas biológicos

fundamentales., abriendo así las puertas a novedosas líneas de investigación y tal vez a la concepción de nuevas aplicaciones biotecnológicas.

En nuestros resultados, los procesos de óxido-reducción se muestran como un componente clave, como lo demuestran los análisis GO y KO. El estrés oxidativo es a menudo capaz de inducir una respuesta transcripcional en los organismos. En *L. polyedrum*, el estrés oxidativo inducido por la exposición a metales dio lugar a aumentos bruscos de la actividad de la enzima de defensa superóxido dismutasa (Walsh *et al.*, 2005), y el incremento de dicha actividad depende del tipo de metal, su concentración y el del tiempo de exposición (Okamoto *et al.*, 1996; Okamoto y Colepicolo, 1998). De igual forma, una micro-matriz de 3.500 genes de *P. lunula* reveló que 204 y 37 genes aumentaron en abundancia de 2 a 4 veces después del tratamiento con nitrito de sodio 1 mM o paraquat 0,5 mM, respectivamente.

En este punto resulta fundamental tener en cuenta que la enzima antioxidante GST (GenBank AAN85429.1) (Okamoto y Hastings, 2003), cuya presencia se ha detectado en nuestros resultados, también posee el dominio pfam05295 (Luciferasa_N). Además del dominio pfam05295 (LCF/LBP N-terminal), también está presente un dominio similar a GST-N-Sigma, que pertenece a la superfamilia tiorredoxina. Funcionan como oxidorreductasas de disulfuro (PDO), que alteran el estado redox de las proteínas diana mediante la oxidación reversible de su sitio activo ditiol. El grupo tiol de la cisteína, en el estado reducido, puede donar un equivalente de reducción a otras moléculas inestables, como las especies reactivas al O₂ (Hartley *et al.*, 1995; Yamauchi *et al.*, 2010), como es el caso de la luciferina.

Se ha reportado que en *P. lunula* el gen que codifica la enzima antioxidante GST tiene una extensión adicional de 255 pb en su extremo (Okamoto y Hasting 2003a), que codifica una secuencia N-terminal previamente encontrada en las proteínas LCF y LBP de *L. polyedrum* (Lee *et al.*, 1993; Bae y Hastings, 1994). La similitud entre los extremos N de GST, LCF y LBP es de alrededor del 45 %, en comparación con el 50 % entre *L. polyedrum* LCF y LBP (Li *et al.*, 1997); se sugirió la recombinación de exones

como una posible explicación para la homología (Li y Hastings, 1998). Aunque la función de esta secuencia N-terminal sigue siendo un enigma, la presencia de esta región conservada en la enzima GST indica que su función no está restringida a la bioluminiscencia y podría estar relacionada con algún otro papel, como el procesamiento celular o la localización de estas proteínas. También es importante señalar que en *P. lunula*, a diferencia de otros dinoflagelados bioluminiscentes, las cantidades de luciferina y LCF son constantes a lo largo del día y la noche (Knaust *et al.*, 1998); así, en lugar de la síntesis diaria *de novo* y la destrucción de todos los componentes, el ritmo parece implicar cambios en su localización intracelular (Seo y Fritz, 2000; Okamoto *et al.*, 2001). Es probable que la proteína GST esté involucrada en este proceso.

Un estudio con trazadores de aminoácidos ha confirmado la producción intracelular de luciferina en *P. lunula* (Wu *et al.*, 2003). En este sentido, Fresneau y Arrio (1988) argumentan que la luminiscencia de *P. lunula* está controlada por el estado de reducción del precursor de la luciferina. Se informó que los análisis de aminoácidos después de hidrólisis ácida han demostrado que la luciferina, el sustrato de la reacción bioluminiscente y su precursor P630 (llamado así por la longitud de onda máxima de excitación, 630 nm), posee el mismo resto peptídico. Esta molécula (P630) es un cromopéptido más estable que la luciferina en soluciones metanólicas a baja temperatura. P630 está compuesto por un polipéptido (4,8 kDa) y un tetrapírol lineal como la luciferina (600 Da). El peso molecular del cromopéptido resultante sería de 5.400 Da. Los cationes pueden oxidar el P630 o romper el enlace entre la cadena peptídica y el tetrapírol extendido. La reducción de P630 se realiza enzimáticamente mediante una oxido-reductasa dependiente de NAD(P)H o químicamente mediante 2-mercaptoetanol o ditiotreitól. El estado de reducción se controló mediante la absorción y la emisión de fluorescencia, que revelaron un cambio conformacional del cromopéptido en función del pH. P630 en su estado reducido tiene las mismas características espectrales que la luciferina purificada. La luciferasa puede oxidar esta molécula reducida con una emisión de luz a 480 nm. Es muy importante tener en cuenta que la luciferina purificada, en metanol a -20 °C, se convirtió parcialmente en P630. Estas observaciones

sugieren que la inter-conversión de P630-luciferina podría constituir el equilibrio de óxido-reducción. De igual forma, los análisis de aminoácidos después de hidrólisis ácida han demostrado que la luciferina, el sustrato de la reacción bioluminiscente de *P. lunula* y su precursor, P630, poseen la misma huella peptídica (Fresneau y Arrio, 1988).

Estas observaciones sugieren que la reducción de P630 conlleva a la generación de la luciferina. La forma oxidada parece, en estas condiciones, ser el precursor de la luciferina (Fresneau *et al.*, 1986). Según Fresneau y Arrio (1988), la emisión de luz en *P. lunula* es un proceso complejo controlado por al menos dos reacciones sucesivas (Fig. 19). El primero es la reducción del precursor de luciferina P630 por una reductasa dependiente de NAD(P)H (Fresneau *et al.*, 1986). La segunda es la conocida reacción de LCF-luciferina. Dado que esta molécula se reduce reversiblemente, puede considerarse como un punto de intercambio de poder reductor que involucra una vía de transferencia de electrones actualmente desconocida que incluye NAD(P) H. En la *P. lunula*, la transferencia metabólica de electrones que regula el nivel reducido de P630 puede considerarse como la fuente de luminiscencia (Fresneau y Arrio, 1988). P630 parece ser una molécula importante involucrada en reacciones complejas moduladas en relación a la luz que están muy difundidas en el metabolismo de las plantas (Buchanan, 1985).

En este punto es muy importante tener en cuenta que Nakamura *et al.* (1989) también informaron la formación de un producto con un color azul profundo característico durante el proceso de purificación de luciferina de dinoflagelado. Esta sustancia (llamada en este caso compuesto azul) fue aislada y purificada. Nakamura *et al.* (1989) reportaron que el producto de oxidación con aire mostró máximos de absorción UV-visible a 633 y 590 nm (hombro), lo que sugiere la presencia de un cromóforo más conjugado que el encontrado en la luciferina. El espectro UV-visible informado para el compuesto azul fue el siguiente: (80% de metanol que contiene 0.1% de NH₄OAc) 234, 254, 315, 370, 410, 590 (hombro) y 633 nm; y el espectro de masa FAB fue (glicerol) m / z 587 [(M - 2Na + 3H) +], 609 [(M - Na + H) '] y 631 [(M + H)']. Estos datos concuerdan con los reportados por Fresneau *et al.* (1986) para

P630: absorción máxima de P630 (oxidado) 630-370-315-250 nm; P630 (reducido) 390 nm; P630 excitación de fluorescencia máxima 630 nm; Por lo tanto, se podría sugerir que el compuesto azul, el precursor P630 y la luciferina son estadios diferentes de la misma molécula (Tabla 3).

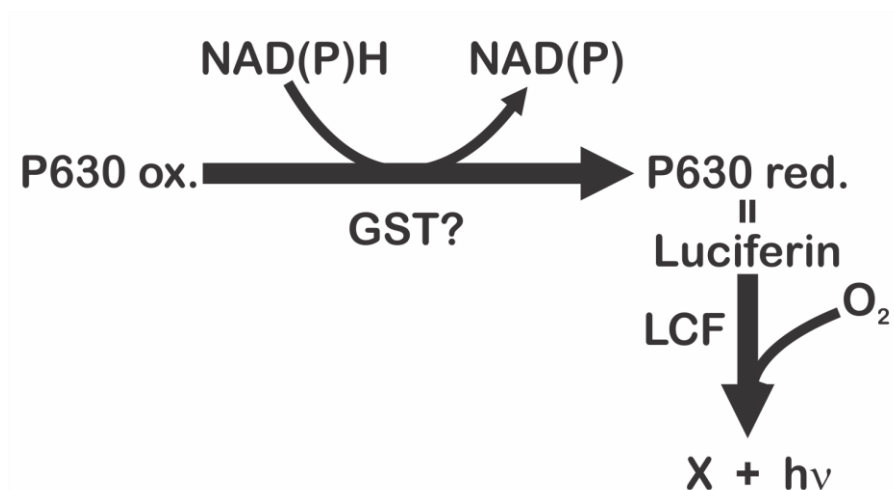


Figura 19. Reacción enzimática acoplada responsable de la bioluminiscencia de *P. lunula* (Fuente: modificado de Fresneau *et al.*, 1986).

Tabla 3. Absorción, excitación de fluorescencia y emisión máxima (nm) de P630 oxidado, P630 reducido y luciferina en 10 mM Hepes-NaOH buffer a pH 7.2 (Fresneau *et al.*, 1986).

Sustratos	Absorción máxima (nm)	Excitación de fluorescencia máxima (nm)	Emisión de fluorescencia máxima (nm)
P630 oxidado	630-370-315-250	630	675
P630 reducido	390	390	480
Luciferina	390-250	390	480

Según Fresneau y Arrio (1988), la luminiscencia se puede considerar como un proceso metabólico involucrado en la regulación del exceso de poder reductor intracelular producido por la fotosíntesis y por la respiración. Este supuesto es consistente con la evidencia de una cloro-respiración específica reportada por Bennoun (1982) para cloroplastos de plantas superiores y otras observaciones en cianobacterias (Pescheck, 1982; Omata y Murata, 1985) de una vía alternativa de respiración en membranas tilacoides fotosintéticas (Doussiere *et al.*, 1979).

Además, los estudios centrados en la desintoxicación dependiente de NADPH de carbonilos reactivos, indican que las células eucariotas también utilizan otros sistemas para hacer frente a los efectos nocivos de los carbonilos reactivos; La vía GST, por ejemplo, que está alimentada por el glutatión (GSH) o el ciclo redox de la tiorredoxina, conjuga los aldehídos con el GSH, lo que proporciona un importante mecanismo de desintoxicación de los carbonilos reactivos. En los animales, la principal vía de desintoxicación es dependiente de GSH, y se ha demostrado que las vías alcohol / aldehído deshidrogenasa y GST representan el 10 y el 55 % de la eliminación de aldehído por los hepatocitos, respectivamente (Hartley *et al.*, 1995). Un sistema de desintoxicación dependiente de GSH también es funcional en las plantas. Las proteínas GST constituyen una gran familia; Algunos de los miembros están localizados en los cloroplastos y las mitocondrias. Debido a que los cloroplastos contienen un orden milimolar de GSH, se cree que el sistema de desintoxicación dependiente de GSH es sustancialmente efectivo (Yamauchi *et al.*, 2011).

Las GST, anteriormente conocidas como ligandinas, comprenden una familia de isozimas metabólicas eucarióticas y procarióticas fase II, mejor conocidas por su capacidad para catalizar la conjugación de la forma reducida de GSH a sustratos xenobióticos con el propósito de desintoxicación (Allocati *et al.*, 2009; Atkinson *et al.*, 2009). ¿Podría la GST ser la enzima reductasa NAD(P)H-dependiente que controla el estado de reducción de P630? La enzima GST también está asociada con el citocromo P450. La presencia de esta GST, con un dominio característico del sistema bioluminiscente (pfam05295), que forma parte del ciclo metabólico del citocromo P450, podría indicar que la enzima GST está involucrada en el proceso de síntesis y/o almacenamiento de luciferina a través de un sistema de transferencia de electrones desconocido hasta ahora.

Recientemente, Wang y Liu (2017) sugirieron un mecanismo de catálisis LCF en dinoflagelados vinculados a una transferencia de energía Dexter. Según estos autores, el hecho de que la luciferina, frente a la oxiluciferina, es fluorescente, con un λ_{em} muy similar al máximo de emisión de

bioluminiscencia, se confirmó utilizando la teoría funcional de la densidad dependiente del tiempo (TD-DFT). En base a esto, se propuso que un intermedio de oxiluciferina en estado excitado formado durante la catálisis LCF podría transferir energía a una segunda molécula de luciferina, o un análogo, que luego serviría como bioluminóforo al relajarse con la emisión radiactiva de luz.

Sin embargo; de acuerdo con Ngo y Mansoorabadi (2017), el proceso implica un mecanismo de luminiscencia de intercambio de electrones iniciado químicamente (CIEEL, por sus siglas en inglés) y/o una reacción de transferencia de carga intramolecular retorcida (TICT, por sus siglas en inglés). Estos autores sugirieron que, en los dinoflagelados, un intermediario en estado excitado derivado de la reacción entre O_2 y luciferina, del tipo gem-diol(ato), puede funcionar directamente como luminóforo. Ngo y Mansoorabadi (2017) también utilizaron el TD-DFT para estudiar los diversos mecanismos nominados de catálisis LCF. La comparación entre la longitud de onda de emisión y la viabilidad termodinámica, sugiere un gem-diol(ato) intermedio como el bioluminóforo, sobre un hidroperóxido o anión peroxi. Estos autores indican que, si el ciclo catalítico de LCF comienza con el isómero E de la luciferina, que es más estable, el proceso probablemente implique un mecanismo de luminiscencia de intercambio electrónico (CIEEL) iniciado químicamente. Este proceso se ha referido a explicar otras reacciones bioluminiscentes, como en el caso de las luciérnagas (Koo *et al.*, 1978; Wilson y Hastings, 1998; Vacher *et al.*, 2018). Sin embargo, si la luciferina tiene la configuración Z, los datos sugieren que un intermediario de gem-diol(ato) en estado excitado retorcido podría servir como bioluminóforo y, por lo tanto, LCF catalizaría una reacción de transferencia de carga intramolecular retorcida (TICT). Los estados de TICT se han relacionado con varias reacciones fotoquímicas; sin embargo, hasta hace poco no se habían relacionado con ningún sistema bioluminiscente (Grabowski *et al.*, 2003; Ngo y Mansoorabadi 2017). ¿Podría el compuesto azul - P630 ser considerado el intermediario gem-diol(ato) relacionado con este proceso? Esta es una pregunta que requiere más investigación.

Una búsqueda exhaustiva en las bases de datos de GenBank reveló que la presencia del dominio pfam05295 (Luciferase_N) en el GST solo se ha informado para *P. lunula* (Okamoto y Hastings 2003; Gornik *et al.*, 2015; Fajardo *et al.*, 2019). Si esta isoforma está involucrada en el proceso de control de la degradación oxidativa de la luciferina en esta especie, esto podría explicar el hecho de que *P. lunula* contiene más luciferina que cualquier otro dinoflagelado conocido hasta la fecha, hasta 100 veces más que en *L. polyedrum* por ejemplo (Knaust *et al.*, 1998).

En cualquier caso, la hipótesis de que el origen de la luciferina es la clorofila foto-oxidada (Topalov y Kishi, 2001) solo sería plausible para *P. lunula*, que mantiene su luciferina a lo largo del ciclo luz-oscuridad. La luciferina fluorescente solo aparece en *L. polyedrum* al comienzo de la fase oscura (Akimoto *et al.*, 2004), por lo que su síntesis no puede explicarse por el mecanismo de fotooxidación. Por lo tanto, es probable que más de un mecanismo sea responsable de la producción de luciferina, incluso en especies fotosintéticas con LCF estrechamente relacionados (Valiadi e Iglesias, 2013).

4.3 ESTUDIO DE POSIBLES APLICACIONES BIOTECNOLÓGICAS

Existen varias aplicaciones biotecnológicas derivadas del sistema bioluminiscente de *P. lunula*. La bioluminiscencia de esta especie se puede cuantificar por espectrofotometría y relacionarse de manera inversamente proporcional con el aumento de las concentraciones de tóxicos en un medio acuoso. Ejemplos de esta aplicación son los bioensayos Lumitox y QwikLite™ 200, dos kits de prueba portátiles diseñados para medir la toxicidad de contaminantes en ambientes marinos. Más recientemente, se ha desarrollado un bioensayo para evaluar posibles casos de contaminación del agua subterránea, relacionado con la extracción de gas natural y la actividad de perforación no convencional (Craig *et al.*, 2003; Stauber *et al.*, 2008; Hildenbrand *et al.*, 2015).

Existen varias aproximaciones que permitirían el desarrollo de posibles aplicaciones biotecnológicas relacionadas con algunas de las proteínas

identificadas en el proteoma de *P. lunula*. La bioluminiscencia se ha establecido como un fenómeno que se puede ajustar para diseñar ensayos rápidos y muy simples para análisis bioquímicos celulares. Ampliamente popularizada para su uso como reporteros genéticos, la bioluminiscencia se ha vuelto regularmente útil para una amplia gama de metodologías de ensayos, los cuales son reconocidos por su precisión cuantitativa y por la baja interferencia con la muestra (Faysal *et al.*, 2019).

4.3.1 Luciferasa (LCF)

Las LCFs se utilizan ampliamente en biología molecular, biología celular y biotecnología, análisis microscópico y también como gen reportero. En la reacción bioluminiscente, las LCFs requieren de la adición de luciferina pero no así de una fuente de luz externa, como es el caso de las proteínas fluorescentes. Es por ello que la LCF y sus genes se han vuelto muy útiles para propósitos de investigación y también para una variedad de propósitos comerciales (Faysal *et al.*, 2019).

Mediante el uso de la ingeniería genética, la enzima LCF puede producirse en el laboratorio para diferentes propósitos. En primer lugar, clonando los genes de la LCF que luego se transfectará en diferentes organismos o células. Actualmente existen una variedad de organismos como ratones, gusanos de seda y papas que se utilizan para producir esta proteína (Contag y Bachmann, 2002). Cuando la LCF cataliza el sustrato de luciferina adecuado, se emite luz durante la reacción. Equipos sensibles a la luz, como el luminómetro o el microscopio óptico modificado, pueden detectar los fotones que se emiten a partir de la reacción de la LCF para la observación de sistemas o procesos biológicos. La LCF se ha usado principalmente como reportero para evaluar el movimiento transcripcional en células que se transfectan con un material genético específico que contiene el gen de la LCF bajo el control de un promotor de interés. También se utiliza para determinar otras actividades enzimáticas, como es el caso de la actividad de la caspasa, del CYP450, entre otras (Fan y Wood, 2007; Faysal *et al.*, 2019).

En los ensayos de actividad de la quinasa o en los ensayos de viabilidad celular, las LCFs son capaces de identificar la concentración celular de ATP (Fan y Wood, 2007; Meisenheimer *et al.*, 2008). La LCF actúa como una proteína sensor de ATP por biotilación, luego la LCF se inmoviliza en la superficie celular al unirse a un complejo de estreptavidina-biotina que permitirá a la LCF identificar el flujo de salida de ATP de la célula y mostrará con éxito la liberación en tiempo real de ATP durante la reacción de bioluminiscencia (Nakamura *et al.*, 2006). La LCF también puede ser modificada para ser más susceptible a la concentración de ATP al elevar la intensidad de luminiscencia como producto de la alteración de la secuencia de aminoácidos de la proteína (Fujii *et al.*, 2007). Se ha reportado un ensayo muy rápido y sensible que se basa principalmente en el método convencional de indicador doble de LCF que se puede usar como una técnica para diferenciar la sección más pequeña de la región promotora de un gen. Este método es muy conveniente para estudiar tanto promotores constitutivos (telomerasa) como inducibles (dependientes de NF-kappaB) (Alcaraz-Pérez *et al.*, 2008).

El uso de la bioluminiscencia constituye una técnica muy poderosa para la generación de imágenes en organismos completos (tanto en imágenes *in vivo* como *in vitro*) para el análisis de poblaciones celulares en animales vivos como ratones (Greer y Szalay, 2002). Hay muchas células (por ejemplo, células T, células madre de médula ósea, células cancerosas) que pueden modificarse para que la LCF permita su visualización no invasiva dentro de las células de animales vivos por medio de cámaras y dispositivo especiales y que se ha utilizado para recopilar valiosa información referente a la carcinogénesis y la respuesta de los tumores en diferentes modelos animales (Lyons *et al.*, 2003; Becher y Holland, 2006). La intensidad de la señal se puede calcular mediante un método de imágenes *in vivo* que puede depender de una variedad de factores, como la tasa de incorporación de luciferina a través del peritoneo, la circulación sanguínea, la permeabilidad de la membrana celular, la disponibilidad de cofactores, el pH intracelular, de la naturaleza del tejido subyacente, así como de la cantidad de la LCF (Inoue *et al.*, 2006). Como la LCF es una proteína susceptible al calor, lo que es de interés y aplicable para estudiar la desnaturalización de las proteínas para verificar la capacidad

defensiva de las proteínas de choque térmico, por lo tanto existen muchas posibilidades para la aplicación práctica de la LCF (Massoud *et al.*, 2007).

Por ejemplo, en el campo de los ensayos de expresión génica, destaca la posibilidad de desarrollar plásmidos con la fusión del gen de la proteína LCF. Esta aproximación pudiese explotar la posibilidad de desarrollar un análisis de corroboración de la expresión de una proteína diana fusionada al gen de la LCF de *P. lunula*, y más aún podría ser utilizada en análisis simultáneos de expresión de varias proteínas si se acopla la reacción al uso de diferentes LCFs, por ejemplo, de luciérnaga y/o de cnidario (cypridina); permitiendo así la monitorización de un análisis dual, o de triple reportero (Suzuki *et al.*, 2005).

4.3.2 Glutación-S-transferasa (GST)

En el caso de la proteína GST, existe la posibilidad de generar etiquetas GST y ensayos despleables basados en la GST de *P. lunula*. Se puede fusionar la GST a una proteína de interés para purificarla de la solución en un proceso conocido como un ensayo de extracción. Esto se logra insertando la secuencia de codificación de ADN de GST junto a la que codifica la proteína de interés. Así, después de la transcripción y traducción, la proteína GST y la proteína de interés se expresarán juntas como una proteína de fusión. Debido a que la proteína GST tiene una fuerte afinidad de unión por GSH, las perlas recubiertas con el compuesto se pueden agregar a la mezcla de proteínas; como resultado, la proteína de interés unida a la GST se ligará a las perlas, aislando la proteína del resto de las que están en solución. Las perlas se recuperan y se lavan para separar la proteína de interés, dando como resultado una proteína purificada. Esta técnica se puede utilizar para dilucidar las interacciones directas proteína-proteína (Benard y Bokoch, 2002; Ren *et al.*, 2003). Una etiqueta GST se usa a menudo para separar y purificar proteínas que contienen la proteína de fusión GST. La etiqueta tiene un tamaño de 220 aminoácidos (aproximadamente 26 KDa) (Long *et al.*, 2011). Esta etiqueta GST puede fusionarse con el extremo N o C de una proteína. Adicionalmente, muchos plásmidos etiquetados con GST comercialmente disponibles incluyen

un dominio de trombina para la escisión de la etiqueta GST durante la purificación de proteínas (Benard y Bokoch, 2002; Tinta *et al.*, 2012).

Por otro lado, las GSTs están relacionadas con los mecanismos de eliminación y el metabolismo de xenobióticos. Este término se refiere a una sustancia química que se ubica en el interior de un organismo y que no se ha originado de forma natural. También se incluyen bajo esta denominación sustancias encontradas en concentraciones mucho más elevadas a lo normal. Algunas sustancias naturales pueden convertirse en xenobióticos cuando son absorbidos por un organismo diferente al que la produjo, como por ejemplo la absorción de hormonas humanas por parte de los peces que se encuentran aguas abajo de plantas de tratamiento, o las defensas químicas sintetizadas por algunos organismos como protección contra la depredación (Mansuy, 2013). Esta denominación se usa generalmente para referirse a sustancias como las dioxinas y los bifenilos policlorados, ya que se interpreta que las mismas son ajenas a cualquier sistema biológico, es decir, son sintetizadas artificialmente por el hombre. Se pueden clasificar como pesticidas, contaminantes ambientales, carcinógenos, hidrocarburos y medicamentos (Singleton, 1994).

Estas sustancias se eliminan a través del metabolismo de xenobióticos. En mamíferos, el proceso engloba procesos de desactivación y excreción, y tiene lugar principalmente en el hígado. Las heces, la orina, el sudor y la respiración son las principales vías de excreción. Algunas enzimas hepáticas se encargan del metabolismo de los xenobióticos, primero promoviendo su activación mediante procesos de reducción, oxidación, hidratación e hidrólisis; y luego conjugando el metabolito secundario activo con GSH o ácido glucurónico, a lo que siguen procesos de excreción mediante la orina o la bilis. Un ejemplo de enzimas relacionadas con el metabolismo xenobiótico lo constituye el citocromo P450 microsomal hepático. Las proteínas que participan en los procesos de metabolización de xenobióticos son muy importantes para el sector farmacéutico, ya que son responsables de la descomposición de los medicamentos (Lu *et al.*, 2015).

4.3.3 Transferencia de Carga Intramolecular Retorcida (TICT)

El estudio del mecanismo TICT puede ser explotado para el diseño de estrategias orientadas al desarrollo de novedosas moléculas funcionales. Por lo tanto, la fluorescencia de TICT es muy prometedora en aplicaciones como diodos orgánicos emisores de luz (OLED), quimiosensores y dispositivos foto-voltáicos. Recientemente se han propuesto algunos de los avances en la investigación de TICT desde el punto de vista de la química de los materiales y las aplicaciones potenciales en la investigación de ciencias de la vida, diagnóstico e imágenes biomédicas, en dispositivos optoelectrónicos y aplicaciones foto-voltáicas (Sasaki *et al.*, 2016).

La transferencia de carga intramolecular retorcida (TICT, por sus siglas en inglés) es un proceso de transferencia de electrones que se produce durante la foto-excitación en moléculas que generalmente consisten en una parte donante y aceptora unida por un solo enlace. Después de la torsión intramolecular, el estado TICT regresa al estado fundamental ya sea a través de la emisión de desplazamiento o por la relajación no radiactiva. Las propiedades de emisión son potencialmente dependientes del medio ambiente, lo que hace que los fluoróforos basados en TICT sean sensores ideales para la detección de solventes, micro-viscosidades y especies químicas. Recientemente, se ha descubierto que varios materiales basados en TICT se vuelven fluorescentes tras la agregación. Además, varios estudios recientes en optoelectrónica orgánica, óptica no lineal y conversiones de energía solar utilizaron el concepto de TICT para modular la mezcla de estado electrónico y el acoplamiento en los estados de transferencia de carga. Los ejemplos anteriores ilustran cómo investigadores de diversas áreas del conocimiento han utilizado el fenómeno TICT con fines que van desde los análisis con reporteros y la bio-imagen, a las técnicas de recolección de energía y visualización / iluminación. Estos ejemplos también ilustran la fortaleza de TICT, es decir, su capacidad de ajuste cuando los grupos de donantes, aceptadores y colgantes están diseñados ingeniosamente. Dicha química orientada a la aplicación solo se puede realizar bien si se entienden suficientemente las propiedades foto-físicas básicas de la reacción TICT, así como su relación con la estructura molecular. Por lo tanto, es

importante que los investigadores del TICT, tanto de la ciencia de los materiales como de la biología y fisicoquímica, cooperen para llenar estos vacíos en beneficio del desarrollo de nuevos fluoróforos (Sasaki *et al.*, 2016).

4.4 PERSPECTIVAS FUTURAS

Todos estos hallazgos toman aun más valor si se consideran bajo la perspectiva de la biología sintética y de la ingeniería genética, ya que las microalgas se han erigido como una de las biorefinerías modelo, y que prometen desarrollar aun más su potencial en los años por venir. Siguiendo este marco conceptual, y basados en estas plataformas biológicas, se pueden diseñar cepas que aprovechen las ventajas de manipulación de diversos genes. Ejemplo de ello son los reportes exitosos en los cuales a través de la implementación de técnicas de ingeniería genética o metabólica se ha logrado el aumento de la productividad de ciertas cepas para la producción de compuestos de alto valor añadido. Es así como toda la información generada en este estudio podrá ser utilizada como un marco conceptual a partir del cual concebir, diseñar o complementar nuevas líneas de investigación.

Por ejemplo, mediante la búsqueda y caracterización de la LBP y GST (isoforma con la presencia del dominio pfam05295 Luciferase_N) en otras especies pertenecientes al género *Pyrocystis* (*P. noctiluca* y *P. fusiformis*). Cabe destacar que hasta ahora se pensaba que todos los miembros de este género no presentaban a la LBP como un componente de su maquinaria bioluminiscente, por lo que su identificación en dichas especies también resultaría un nuevo reporte. La combinación sinérgica entre la data generada en este estudio y la implementación de las técnicas de manipulación e ingeniería genética disponibles actualmente en el caso de las microalgas podría brindar también interesantes perspectivas. Profundizando aún más, se podría aplicar la técnica de ARNi (ARN de interferencia) o CRISPRi para silenciar la expresión de la GST (con dominio pfam05295 Luciferase_N) en *P. lunula* y así poder estudiar cuál sería el efecto de esta inhibición sobre el sistema bioluminiscente de estos organismos.

El ARN de interferencia (ARNi) es una herramienta prometedora para el silenciamiento selectivo de genes en algas que albergan la maquinaria ARNi funcional. En el caso de las microalgas, el ARNi como herramienta de genética inversa se ha reportado solo en unas pocas especies (Liang *et al.*, 2018). El primer informe fue en *C. reinhardtii*, donde un sistema que consiste en transgenes de repeticiones invertidas en tándem que activaron sistemáticamente el silenciamiento simultáneo de un gen con un fenotipo seleccionable inducido por ARNi. (Rohr *et al.*, 2004). CRISPRi se ha aplicado recientemente en ingeniería metabólica para manipular la regulación de genes en ciertas rutas relacionadas con la síntesis de productos de interés (Elhadi *et al.*, 2016; Gordon *et al.*, 2016; Huang *et al.*, 2016; Li *et al.*, 2016; Lv *et al.*, 2015).

Como prueba de concepto, algunos estudios han demostrado que el uso del sistema CRISPR-Cas9 controlado por vectores en especies de microalgas como *C. reinhardtii*, *P. tricornutum* y *Nannochloropsis sp.* es factible (Jiang *et al.*, 2014; Nymark *et al.*, 2016; Wang *et al.*, 2016). Debido a que CRISPRi logró una regulación génica sin roturas en la doble cadena de ADN, los transformantes no tienen los problemas de las técnicas de edición de genes tradicionales, como la incorporación aleatoria (Fajardo *et al.*, 2019).

5. CONCLUSIONES

1. Los resultados obtenidos ponen por primera vez a disposición de la comunidad científica un transcriptoma y proteoma de referencia de la especie *P. lunula*, así como las secuencias de 3.182 proteínas, incluidas 175 proteínas no identificadas hasta ahora, que se han depositado en las bases de datos del NCBI SRA y ProteomeXchange, respectivamente.
2. Se ha demostrado la presencia de la proteína LBP en *P. lunula*. Además, se han detectado las isoformas LBP y LCF/LBP, lo que indica la importancia de estas proteínas en el mecanismo de bioluminiscencia de esta especie.
3. Se destaca la presencia de una isoforma de la proteína GST (con el dominio pfam05295 Luciferase_N), que podría estar involucrada en el proceso de control de la degradación oxidativa de la luciferina.
4. Este transcriptoma y proteoma ayudarán a acelerar la investigación genómica funcional en esta especie y, quizás, en otras microalgas importantes desde el punto de vista comercial y ambiental.
5. La combinación sinérgica entre el análisis transcriptómico y proteómico constituye una aproximación efectiva para producir secuencias de alta calidad en especies que, debido a la complejidad de sus genomas, no se han secuenciado.
6. Existen diversas posibilidades para el desarrollo de aplicaciones biotecnológicas relacionadas con algunas de las proteínas identificadas en el proteoma de *P. lunula*: i) la proteína LCF, que puede ser usada para el desarrollo de ensayos simultáneos de múltiples genes reporteros; y ii) la GST, para el desarrollo de técnicas de aislamiento y purificación de proteínas, o para su uso por la industria farmacéutica.

5. CONCLUSIONS

1. The results obtained make for the first time available to the scientific community a transcriptome and proteome of reference of the *P. lunula* species, as well as the sequences of 3,182 proteins, including 175 proteins not identified so far, which have been deposited in the NCBI SRA and ProteomeXchange databases, respectively.
2. The presence of LBP protein in *P. lunula* has been demonstrated. In addition, LBP and LCF / LBP isoforms have been detected, indicating the importance of these proteins in the bioluminescence mechanism of this species.
3. The presence of an isoform of the GST protein is highlighted (with the domain pfam05295 Luciferase_N), which could be involved in the process of controlling the oxidative degradation of luciferin.
4. This transcriptome and proteome will help accelerate functional genomic research in this species and, perhaps, in other important microalgae commercially and environmentally.
5. The synergistic combination between transcriptomic and proteomic analysis constitutes an effective approach to produce high quality sequences in species that, due to the complexity of their genomes, have not been sequenced.
6. There are several possibilities for the development of biotechnological applications related to some of the proteins identified in the *P. lunula* proteome: i) LCF protein, which can be used for the development of simultaneous tests of multiple reporter genes; and ii) GST, for the development of protein isolation and purification techniques, or for use by the pharmaceutical industry.

6. REFERENCIAS

1. Abrahams M, Townsend L (1993) Bioluminescence in dinoflagellates: A test of the burglar alarm hypothesis. *Ecology* 74: 258-260.
2. Akimoto H, Wu C, Kinumi T, Ohmiya Y (2004) Biological rhythmicity in expressed proteins of the marine dinoflagellate *Lingulodinium polyedrum* demonstrated by chronological proteomics. *Biochemical and Biophysical Research Communications* 315: 306-312.
3. Alcaraz-Pérez F, Mulero V, Cayuela M (2008) Application of the dual-luciferase reporter assay to the analysis of promoter activity in Zebrafish embryos. *BMC Biotechnology* 8: 81.
4. Allocati N, Federici L, Masulli M, Di Ilio C (2009) Glutathione transferases in bacteria. *FEBS Journal* 276(1): 58-75.
5. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-3402.
6. Anand V, Singh P, Banerjee C, Shukla P (2017) Proteomic approaches in microalgae: Perspectives and applications. *Biotech* 7: 197.
7. Andrianantoandro E, Basu S, Karig D, Weiss R (2006) Synthetic biology: New engineering rules for an emerging discipline. *Molecular Systems Biology* 2: 1–14.
8. Armbrust E, Berges J, Bowler C, Green B, Martinez D, *et al.* (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306: 79–86.
9. Atkinson H, Babbitt P (2009) Glutathione transferases are structural and functional outliers in the thioredoxin fold. *Biochemistry* 48(46): 11108-11116.
10. Bachvaroff T, Place A (2008). From stop to start: Tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate *Amphidinium carterae*. *PLoS One* 3: e2929.
11. Bae Y, Hastings J (1994) Cloning, sequencing and expression of dinoflagellate luciferase DNA from a marine alga, *Gonyaulax polyedra*. *Biochimica et Biophysica Acta* 1219: 449-456.
12. Banerjee C, Singh K, Shukla P (2016) Microalgal bioengineering for sustainable energy development: Recent transgenesis and metabolic engineering strategies. *Biotechnology Journal* 11: 303–314.
13. Becher O, Holland E (2006) Genetically engineered models have advantages over xenografts for preclinical studies. *Cancer Research* 66: 3358-3359.
14. Beisser D, Kaschani F, Graupner N, Grossmann L, Jensen M, Ninck S, *et al.* (2017) Quantitative proteomics reveals ecophysiological effects of light and silver stress on the mixotrophic protist *Poteroiochromonas malhamensis*. *PLoS ONE* 12 (1): e0168183.
15. Ben F, Elleuch F, Ben H, Garnier M, Saint-Jean B, Barkallah M, *et al.* (2017) Proteomic analysis of the chlorophyte *Dunaliella* new strain AL-1 revealed global changes of metabolism during high carotenoid production. *Marine Drugs* 15: 293.
16. Benard V, Bokoch G (2002) Assay of Cdc42, Rac, and Rho GTPase activation by affinity methods. *Methods in Enzymology* 345: 349-359.
17. Bennoun P (1982) Evidence for a respiratory chain in the chloroplast *Proceeding of the National Academy of Sciences of the United States of America* 79: 4352-4356.
18. Biggley W, Swift E, Buchanan R, Seliger H (1969) Stimulable and spontaneous bioluminescence in the marine dinoflagellates *Pyrodinium bahamense*, *Gonyaulax polyedra* and *Pyrocystis lunula*. *Journal of General Physiology* 54: 96–122.
19. Blatti L, Michaud J, Burkart M (2013) Engineering fatty acid biosynthesis in microalgae for sustainable biodiesel. *Current Opinion in Chemical Biology* 17: 496–505.
20. Bogen C, Al-Dilaimi A, Albersmeier A, Wichmann J, Grundmann M, Rupp O, *et al.* (2013) Reconstruction of the lipid metabolism for the microalga *Monoraphidium neglectum* from its genome sequence reveals characteristics suitable for biofuel production. *BMC Genomics* 14: 926.
21. Boisvert S, Laviolette F, Corbeil J (2010) Ray: Simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *Journal of Computational Biology* 17: 1519-1533.
22. Bowler C, Allen A, Badger J, Grimwood J, Jabbari K, *et al.* (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456: 239–244.
23. Buchanan B (1985) *Thioredoxin system and glutaredoxin systems*. Raven Press, New York.
24. Burkenroad M (1943) A possible function of bioluminescence. *Journal of Marine Research* 5: 161-164.

25. Burki F, Keeling P (2014) Rhizaria. *Current Biology* 24: 103–107.
26. Buskey E, Mann C, Swift E (1987) Photophobic responses of calanoid copepods: Possible adaptive value. *Journal of Plankton Research* 9: 857–870.
27. Buskey E, Mills L, Swift E (1983) The effects of dinoflagellate bioluminescence on the swimming behavior of a marine copepod. *Limnology and Oceanography* 28: 575–579.
28. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10: 421.
29. Carrasco R, Fajardo C, Guarnizo P, Vallejo R, Fernandez-Acero F (2018) Biotechnology applications of microalgae in the context of EU “blue growth” initiatives. *Journal of Microbiology and Genetics* 10: 2574–7371
30. Chan L, Hodgkiss I, Lu S, Lo S (2004) Use of two-dimensional gel electrophoresis proteome reference maps of dinoflagellates for species recognition of causative agents of harmful algal blooms. *Proteomics* doi:10.1002/pmic.200300548.
31. Chan Y, Wong J (2007) Concentration-dependent organization of DNA by the dinoflagellate histone-like protein HCc3. *Nucleic Acids Research* 35: 2573–2583.
32. Chevreux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Journal of Computer and System Sciences* 99: 45–56.
33. Chisti Y (2007) Biodiesel from microalgae. *Biotechnology Advances* 25: 294–306.
34. Cock J, Sterck L, Rouz P, Scornet D, Allen A, Amoutzias G, *et al.* (2010) The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature* 465: 617–621.
35. Colepiccolo P, Roenneberg T, Morse D, Taylor W, Hastings J (1993) Circadian regulation of bioluminescence in the dinoflagellate *Pyrocystis lunula*. *Journal of Phycology* 29: 173–179.
36. Conesa A, Gotz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008: 619832.
37. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
38. Contag C, Bachmann M (2002) Advances *in vivo* bioluminescence imaging of gene expression. *Annual Review of Biomedical Engineering* 4: 235–260.
39. Costas E, Goyanes V (2005) Architecture and evolution of dinoflagellate chromosomes: An enigmatic origin. *Cytogenetics and Genome Research* 109: 268–275.
40. Craig J, Klerks P, Heimann K, Waits J (2003) Effects of salinity, pH and temperature on the re-establishment of bioluminescence and copper or SDS toxicity in the marine dinoflagellate *Pyrocystis lunula* using bioluminescence as an endpoint. *Environmental Pollution* 125: 267–275
41. Craig J, Priede I (2012) Bioluminescence: a source of marine energy? *The Crown Estate*, pp. 27.
42. Croft T, Moulin M, Webb E, Smith G (2007) Thiamine biosynthesis in algae is regulated by riboswitches. *Proceedings of the National Academy of Sciences of United States Of America* 104: 20770–20775.
43. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Research* 14: 1188–1190.
44. Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, *et al.* (2012) Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492: 59–65.
45. Daboussi F, Leduc S, Marchal A, Dubois G, Guyot V, Perez-Michaut C *et al.* (2014) Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. *Nature Communications* 5: 3831
46. De Moreno Diaz La Espina S, Alverca E, Cuadrado A, Franca S (2005) Organization of the genome and gene expression in a nuclear environment lacking histones and nucleosomes: The amazing dinoflagellates. *European Journal of Cell Biology* 84: 137–149.
47. Del Campo J, Sieracki ME, Molestina R, Keeling PJ, Massana R, *et al.* (2014) The others: Our biased perspective on eukaryotic genomes. *Trends in Ecology & Evolution* 29: 252–259.
48. Derelle E, Ferraz C, Rombauts S, Rouze P, Worden A, *et al.* (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America* 103: 11647–11652.
49. DeSa R, Hastings J (1968) The characterization of scintillons. Bioluminescent particles from the marine dinoflagellate *Gonyaulax polyedra*. *Journal of General Physiology* 51:

- 105–122.
50. Dodge J (1985) The chromosomes of dinoflagellates. *International Review of Cytology* 94: 5–19.
 51. Donnan PH, Ngo PD, Mansoorabadi SO (2017) Constant pH accelerated molecular dynamics investigation of the pH regulation mechanism of dinoflagellate luciferase. *Biochemistry* 57(3): 295–9.
 52. Doussiere J, Sainsard-Chanet A, Vignais P (1979) The respiratory chain of *P. tetraurelia* in wild-type and mutant Cl1. *Biochimica et Biophysica Acta* 548: 236–252.
 53. Dunlap J, Hastings J (1981) The biological clock in *Gonyaulax* controls luciferase activity by regulating turnover. *Journal of Biological Chemistry* 256: 10509–10518.
 54. Eckert R (1965) Bioelectric control of bioluminescence in the dinoflagellate *Noctiluca*. II. *Asynchronous* flash initiation by a propagated triggering potential. *Science*, 147: 1142–1145.
 55. Edgar R (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
 56. Elhadi D, Lv L, Jiang X, Wu H, Chen G (2016) CRISPRi engineering *E. coli* for morphology diversification. *Metabolic Engineering* 38: 358–369.
 57. Emami K, Hack E, Nelson A, Brain C, Lyne F, Mesbahi E *et al.* (2015) Proteomic-based biotyping reveals hidden diversity within a microalgae culture collection: an example using *Dunaliella*. *Scientific Reports* 5: 10036.
 58. Erdner D, Anderson D (2006) Global transcriptional profiling of the toxic dinoflagellate *Alexandrium fundyense* using massively parallel signature sequencing. *BMC Genomics* 7: 88.
 59. Fajardo C, De Donato M, Carrasco R, Martinez-Rodriguez G, Mancera J, Fernandez-Acero F (2019) Advances and challenges in genetic engineering of microalgae. *Reviews in Aquaculture* doi: 10.1111/raq.12322
 60. Fan F, Wood K (2007) Bioluminescent assays for highthroughput screening. *Assay and Drug Development Technologies* 5: 127–136.
 61. Faysal S, MensahSedro D, Akbar H, Saker S (2019) Structure, enzymatic mechanism of action, applications, advantages, disadvantages and modifications of luciferase enzyme. *Journal of Genetic Engineering and Biotechnology Research* 1(1): 1–6.
 62. Fernandez-Acero F, Amil-Ruiz F, Duran-Peña M, Carrasco R, Fajardo C, Guarnizo P, *et al.* (2018) Valorisation of the microalgae *Nannochloropsis gaditana* biomass by proteomic approach in the context of circular economy. *Journal of Proteomics* <https://doi.org/10.1016/j.jprot.2018.10.015>.
 63. Fogel M, Hastings J (1971) A substrate-binding protein in the *Gonyaulax* bioluminescence reaction. *Archives of Biochemistry and Biophysics* 142: 310–321.
 64. Fresneau C, Arrio B (1988) *Pyrocystis lunula* bioluminescence: Physicochemical characterization of the luciferin precursor. *Archives of Biochemistry and Biophysics* 265(1): 22–27.
 65. Fresneau C, Hill M, Lescure N, Arrio B, Dupaix A, Volfin P (1986) Dinoflagellate luminescence: Purification of a NAD(P)H-dependent reductase and of its substrate. *Archives of Biochemistry and Biophysics* 251(2): 495–503.
 66. Fujii H, Noda K, Asami Y, Kuroda A, Sakata M, *et al.* (2007) Increase in bioluminescence intensity of firefly luciferase using genetic modification. *Analytical Biochemistry* 366: 131–36.
 67. Gao C, Wang Y, Shen Y, Yan D, He X, Dai J, *et al.* (2014) Oil accumulation mechanisms of the oleaginous microalga *Chlorella protothecoides* revealed through its genome, transcriptomes, and proteomes. *BMC Genomics* 15 (1): 582.
 68. Garibay A, Barkla B, Vera-Estrella R, Martinez A, Pantoja O (2017) Membrane proteomic insights into the physiology and taxonomy of an oleaginous green microalga. *Plant Physiology* 173: 390–416.
 69. Garnier M, Carrier G, Rogniaux H, Nicolau E, Bougaran G, Saint-Jean B, *et al.* (2014) Comparative proteomics reveals proteins impacted by nitrogen deprivation in wild-type and high lipid-accumulating mutant strains of *Tisochrysis lutea*. *Journal of Proteomics* 105: 107–120.
 70. Gatamaneni L, Orsat V, Lefsrud M (2018) Valuable bioproducts obtained from microalgal biomass and their commercial applications: A review. *Environmental Engineering Research* 23 (3): 229–241.
 71. Ghazalpour A, Bennett B, Petyuk V, Orozco L, Hagopian R, Mungrue I, *et al.* (2011)

- Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genetics* 7(6): e1001393.
72. Gobler CJ, Berry DL, Dyhrman ST, Wilhelm SW, Salamon A, *et al.* (2011) Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proceedings of the National Academy of Sciences of the United States of America* 108: 4352–4357.
 73. Gordon G, Korosh T, Cameron J, Markley A, Begemann M, Pfleger B (2016) CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium *Synechococcus sp.* strain PCC 7002. *Metabolic Engineering* 38: 170–179.
 74. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
 75. Greer L, Szalay A (2002) Imaging of light emission from the expression of luciferases in living cells and organisms: A review. *Luminescence* 17: 43–74.
 76. Gu W, Li H, Zhao P, Yu R, Pan G, Gao S *et al.*, *et al.* (2015) Quantitative proteomic analysis of thylakoid from two microalgae (*Haematococcus pluvialis* and *Dunaliella salina*) reveals two different high light-responsive strategies. *Scientific Reports* 4: 6661.
 77. Guarnieri M, Nag A, Smolinski S, Darzins A, Seibert M, Pienkos P (2011) Examination of triacylglycerol biosynthetic pathways via *de novo* transcriptomic and proteomic analyses in an unsequenced microalga. *PLoS One* 6 (10): e25851.
 78. Guerrero-Sanchez VM, Maldonado-Alconada AM, Amil-Ruiz F, Verardi A, Jorin-Novo JV, Rey MD (2019) Ion Torrent and Illumina, two complementary RNA-seq platforms for constructing the holm oak (*Quercus ilex*) transcriptome. *PLoS One* 14: e0210356.
 79. Guillard RR, Ryther JH (1962) Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Gran. *Canadian Journal of Microbiology* 8: 229–239.
 80. Gupta K, Shukla P (2015) Advanced technologies for improved expression of recombinant proteins in bacteria: Perspectives and applications. *Critical Reviews in Biotechnology* 36 (6): 1089–1098.
 81. Hackett J, Scheetz T, Yoon H, Soares M, Bonaldo, M, Casavant T, Bhattacharya D (2005) Insights into a dinoflagellate genome through expressed sequence tag analysis. *BMC Genomics* 6: 80.
 82. Haddock S, Moline M, Case J (2010) Bioluminescence in the sea. *Annual Review of Marine Science* 2: 443–493.
 83. Hall M, Unch J, Binkowski B, Valley M, Butler B, *et al.* (2012) Engineered luciferase reporter from a deep-sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chemical Biology* 7(11): 1848–1857.
 84. Hamman J, Seliger H (1972) The mechanical triggering of bioluminescence in marine dinoflagellates: Chemical basis. *Journal of Cellular Physiology* 80: 397–408.
 85. Hanley K, Widder E (2017) Bioluminescence in dinoflagellates: Evidence that the adaptive value of bioluminescence in dinoflagellates is concentration dependent. *Photochemistry and Photobiology* 93(2): 519–530.
 86. Hartley D, Ruth J, Petersen D (1995) The hepatocellular metabolism of 4-hydroxynonenal by alcohol dehydrogenase, aldehyde dehydrogenase and glutathione S-transferase. *Archives of Biochemistry and Biophysics* 16: 197–205.
 87. Hastings J, Morin J (1991) Neural and integrative animal physiology. Prosser C (Ed.) Wiley-Interscience, New York, pp. 131–170.
 88. Hastings JW (2013) Circadian rhythms in dinoflagellates: What is the purpose of synthesis and destruction of proteins?. *Microorganisms* 1: 26–32.
 89. Heimann K, Klerks P, Hasenstein K (2009) Involvement of actin and microtubules in regulation of bioluminescence and translocation of chloroplasts in the dinoflagellate *Pyrocystis lunula*. *Botanica Marina* 52: 170–177.
 90. Herring P (1987) Systematic distribution of bioluminescence in living organisms. *Journal of Bioluminescence and Chemiluminescence* 1: 147.
 91. Herzog M, Von Boletzky S, Soyer M (1984) Ultra- structural and biochemical nuclear aspects of eukaryote classification: Independent evolution of the dinoflagellates as a sister group of the actual eukaryotes. *Origin of Life* 13: 205–215.
 92. Hildenbrand Z, Osorio A, Carlton D, *et al.* (2015) Rapid analysis of eukaryotic bioluminescence to assess potential groundwater contamination events. *Journal of Chemistry* <https://doi.org/10.1155/2015/957608>.

93. Huang X, Weber J, Hinson T, Mathieson A, Minocha S (1996) Transient expression of the GUS reporter gene in the protoplasts and partially digested cells of *Ulva lactuca* (Chlorophyta). *Botanica Marina* 39: 467-474.
94. Inoue Y, Tojo A, Sekine R, Soda Y, Kobayashi S, *et al.* (2006) *In vitro* validation of bioluminescent monitoring of disease progression and therapeutic response in leukaemia model animals. *European Journal of Nuclear Medicine and Molecular Imaging* 33: 557-565.
95. Irrgang A, Weise C, Murugaiyan J, Roesler U (2015) Identification of immunodominant proteins of the microalgae *Prototheca* by proteomic analysis. *New Microbes New Infections* 3: 37–40.
96. Jackson A (2007) Tandem gene arrays in *Trypanosoma brucei*: Comparative phylogenomic analysis of duplicate sequence variation. *BMC Evolutionary Biology* 7: 54.
97. Jaeckisch N, Yang I, Wohlrab S, Glöckner G, Kroymann J, Vogel H, Cembella A, John U (2011) Comparative genomic and transcriptomic characterization of the toxigenic marine dinoflagellate *Alexandrium ostenfeldii*. *PLoS One* 6: e28012.
98. Jagadevan S, Banerjee A, Banerjee C, Guria C, Tiwari R, Baweja M, *et al.* (2018) Recent developments in synthetic biology and metabolic engineering in microalgae towards biofuel production. *Biotechnology for Biofuels* 11: 185.
99. Janouskovec J, Gavelis GS, Burki F, Dinh D, Bachvaroff TR, Gornik SG, Bright KJ, Imanian B, Strom SL, Delwiche CF, *et al.* (2017) Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proceedings of the National Academy of Sciences of the United States of America* 114: E171-E180.
100. Jiang W, Brueggeman A, Horken K, Plucinak T, Weeks D (2014) Successful transient expression of Cas9 and single guide RNA genes in *Chlamydomonas reinhardtii*. *Eukaryotic Cell* 13 (11): 1465-1469.
101. Johnson C, Inoue S, Flint A, Hastings J (1985) Compartmentalization of algal bioluminescence- Autofluorescence of bioluminescent particles in the dinoflagellate *Gonyaulax* as studied with image-intensified video microscopy and flow cytometry. *Journal of Cell Biology* 100: 1435–1446.
102. Kaskova Z, Tsarkovaab A, Yampolsky I (2016) 1001 lights: luciferins, luciferases, their mechanisms of action and applications in chemical analysis, biology and medicine. *Chemical Society Reviews* 45: 6048.
103. Kazamia E, Smith G (2014) Assessing the environmental sustainability of biofuels. *Trends in Plant Science* 19: 615–618.
104. Keeling P, Burki F, Wilcox H, Allam B, Allen E, Amaral-Zettler L, *et al.* (2014) The marine microbial eukaryote transcriptome sequencing project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biology* 12(6): e1001889.
105. Kelley L, Mezulis S, Yates C, Wass M, Sternberg M (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10: 845.
106. Kilian O, Benemann C, Niyogi K, Vick B (2011) High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proceedings of the National Academy of Sciences of United States of America* 108: 21265–21269.
107. Kim Y, Yoo W, Lee S, Lee M (2005) Proteomic analysis of cadmium-induced protein profile alterations from marine alga *Nannochloropsis oculata*. *Ecotoxicology* 14: 589–596.
108. Klein-Marcuschamer D, Chisti Y, Benemann R, Lewis D (2013) A matter of detail: Assessing the true potential of microalgal biofuels. *Biotechnology and Bioengineering* 110: 2317–2322.
109. Knaust R, Urbig T, Li L, Taylor W, Hastings J (1998) The circadian rhythm of bioluminescence in *Pyrocystis* is not due to differences in the amount of luciferase: A comparative study of three bioluminescent marine dinoflagellates *Journal of Phycology* 34(1): 167-172.
110. Koo J, Schmidt S, Schuster G (1978) Bioluminescence of the firefly: Key steps in the formation of the electronically excited state for model systems. *Proceeding of the National Academy of Sciences of the United States of America* 75:30-33.
111. Kopylova E, Noé L, Touzet H (2012) SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28(24): 3211-3217.
112. Krasnow R, Dunlap J, Taylor W, Hastings J, Vetterling W, Gooch V (1980) Circadian spontaneous bioluminescent glow and flashing of *Gonyaulax polyedra*. *Journal of Comparative Physiology* 138: 19–26.
113. Krohn I, Alawi M, Forstner K, Wiegandt A, Burkhardt L, Indenbirken D, *et al.* (2017) Insights

- into microalga and bacteria interactions of selected phycosphere biofilms using metagenomic, transcriptomic, and proteomic approaches. *Frontiers in Microbiology* 8: 1941.
114. Kruse O, Rupprecht J, Bader K, Thomas-Hall S, Schenk M, Finazzi G, *et al.* (2005) Improved photobiological H₂ production in engineered green algal cells. *Journal of Biological Chemistry* 280: 34170–34177.
 115. Kumar S, Kumar P, Sandhu S (2018) A review on microalgae strains, cultivation, harvesting, biodiesel conversion and engine implementation. *Biofuels* <https://doi.org/10.1080/17597269.2018.1457314>.
 116. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33: 1870-1874.
 117. Latz M, Bovard M, Groisman A (2008) Bioluminescent response of individual dinoflagellate cells to hydrodynamic stress measured with millisecond resolution in a microfluidic device. *Journal of Experimental Biology* 211: 2865-2875.
 118. Latz M, Case J, Gran R (1994) Excitation of bioluminescence by laminar fluid shear associated with simple Couette flow. *Limnology and Oceanography* 39: 1424-1439.
 119. Latz M, Lee A (1995) Spontaneous and stimulated bioluminescence of the dinoflagellate *Ceratocorys horrida* (Peridinales). *Journal of Phycology* 31: 120–132.
 120. Lee D, Mittag M, Sczekan S, Morse D, Hastings J (1993) Molecular cloning and genomic organization of a gene for luciferin-binding protein from the dinoflagellate *Gonyaulax polyedra*. *Journal of Biological Chemistry* 268: 8842-8850.
 121. Li H, Shen C, Huang C, Sung L, Wu M, Hu Y (2016) CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. *Metabolic Engineering* 38: 293–302.
 122. Li L, Hastings J (1998) The structure and organization of the luciferase gene in the photosynthetic dinoflagellate *Gonyaulax polyedra*. *Plant Molecular Biology* 36: 275-284.
 123. Li L, Hong R, Hastings J (1997) Three functional luciferase domains in a single polypeptide chain. *Proceeding of the National Academy of Sciences of the United States of America* 94: 8954-8958.
 124. Li L, Liu L, Hong R, Robertson D, Hastings J (2001) N-terminal intramolecularly conserved histidines of three domains in *Gonyaulax* luciferase are responsible for loss of activity in the alkaline region. *Biochemistry* 40: 1844-1849.
 125. Li S, Cao X, Wang Y, Zhu Z, Zhang H, Xue S, *et al.* (2017) A method for microalgae proteomics analysis based on modified filter-aided sample preparation. *Applied Biochemistry and Biotechnology* 183: 923–930.
 126. Liang M, Zhu J, Jiang J (2018) High-value bioproducts from microalgae: strategies and progress. *Critical Reviews in Food Science and Nutrition* <https://doi.org/10.1080/10408398.2018.1455030>.
 127. Lidie K, van Dolah F (2007) Spliced leader RNA-mediated trans-splicing in a dinoflagellate, *Karenia brevis*. *Journal of Eukaryotic Microbiology* 54(5): 427-435.
 128. Lin S (2011) Genomic understanding of dinoflagellates. *Research in Microbiology* 162: 551–569.
 129. Lindström J, Grebner W, Rigby K, Selander E (2017) Effects of predator lipids on dinoflagellate defence mechanisms-increased bioluminescence capacity. *Scientific Reports* 7: 13104.
 130. Liñeiro E, Chiva C, Cantoral JM, Sabido E, Fernandez-Acero FJ (2016) Dataset of the *Botrytis cinerea* phosphoproteome induced by different plant-based elicitors. *Data Brief* 7: 1447-1450.
 131. Liu L, Hastings J (2007) Two different domains of the luciferase gene in the heterotrophic dinoflagellate *Noctiluca scintillans* occur as two separate genes in photosynthetic species. *Proceeding of the National Academy of Sciences of the United States of America* 104: 696-701.
 132. Liu L, Wilson T, Hastings J (2004) Molecular evolution of dinoflagellate luciferases, enzymes with three catalytic domains in a single polypeptide. *Proceedings of the National Academy of Sciences of the United States of America* 101: 16555-16560.
 133. Long F, Cho W, Ishii Y (2011) Expression and purification of 15N- and 13C-isotope labeled 40-residue human Alzheimer's β -amyloid peptide for NMR-based structural analysis. *Protein Expression and Purification* 79(1): 16-24.
 134. Longworth J, Wu D, Huete-Ortega M, Wright P, Vaidyanathan S (2016) Proteome response of *Phaeodactylum tricornutum*, during lipid accumulation induced by nitrogen depletion.

- Algal Research 18: 213–224.
135. Lozano J, Schatt P, Botebol H, Verge V, Lesuisse E, Blain S, *et al.* (2014) Efficient gene targeting and removal of foreign DNA by homologous recombination in the picoeukaryote *Ostreococcus*. *The Plant Journal* 78: 1073–1083.
 136. Lu K, Mahbub R, Fox J (2015) Xenobiotics: Interaction with the Intestinal Microflora. *ILAR Journal* 56(2): 218–227.
 137. Lv L, Ren Y, Chen J, Wu Q, Chen G (2015) Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: controllable P(3HB-co-4HB) biosynthesis. *Metabolic Engineering* 29: 160–168.
 138. Lyons SK, Meuwissen R, Krimpenfort P, Berns A (2003) The generation of a conditional reporter that enables bioluminescence imaging of Cre/loxP-dependent tumorigenesis in mice. *Cancer Research* 63: 7042–7046.
 139. Ma Q, Wang J, Lu S, Lv Y, Yuan Y (2013) Quantitative proteomic profiling reveals photosynthesis responsible for inoculum size dependent variation in *Chlorella sorokiniana*. *Biotechnology and Bioengineering* 110: 773–784.
 140. Machabee S, Wall L, Morse D (1994) Expression and genomic organization of a dinoflagellate gene family. *Plant Molecular Biology* 25: 23–31.
 141. Maldonado E, Latz M (2007) Shear-stress dependence of dinoflagellate bioluminescence. *The Biological Bulletin* 212: 242–249.
 142. Mansuy D (2013) Metabolism of xenobiotics: beneficial and adverse effects. *Biologie Aujourd'hui* 207(1): 33–37.
 143. Marcinko C, Martin A, Allen J (2014) Modelling dinoflagellates as an approach to the seasonal forecasting of bioluminescence in the North Atlantic. *Journal of Marine Systems* 139: 261–275.
 144. Marcinko C, Painter S, Martin A, Allen J (2013) A review of the measurement and modelling of dinoflagellate bioluminescence. *Progress in Oceanography* 109: 117–129.
 145. Massoud T, Paulmurugan R, De A, Ray P, Gambhir S (2007) Reporter gene imaging of protein-protein interactions in living subjects. *Current Opinion in Biotechnology* 18: 31–37.
 146. Meisenheimer P, O'Brien M, Cali J (2008) Luminogenic enzyme substrates: The basis for a new paradigm in assay design. *Promega Notes* 100: 22–26.
 147. Moline M, Oliver M, Mobley C, Sundman L, Bensky T, Bergmann T, Bissett W, Case J, Raymond E, Schofield O (2007) Bioluminescence in a complex coastal environment: Temporal dynamics of nighttime water-leaving radiance. *Journal of Geophysical Research: Oceans* 112: C11016
 148. Morishita H, Ohashi S, Oku T, Nakajima Y, Kojima S, Ryufuku M, Ohmiya Y (2002) Cloning and characterization of an active fragment of luciferase from a luminescent marine alga, *Pyrocystis lunula*. *Photochemistry and Photobiology* 75(3): 311–315.
 149. Morse D, Pappenheimer A, Hastings J (1989) Role of a luciferin-binding protein in the circadian bioluminescent reaction of *Gonyaulax polyedra*. *Journal of Biological Chemistry* 264: 11822–11826.
 150. Moses T, Mehrshahi P, Smith A, Goossens A (2017) Synthetic biology approaches for the production of plant metabolites in unicellular organisms. *Journal of Experimental Botany* 15: 4057–4074.
 151. Murugaiyan J, Eravci M, Weise C, Roesler U (2016) Label-free quantitative proteomic analysis of harmless and pathogenic strains of infectious microalgae, *Prototheca spp.* *International Journal of Molecular Sciences* 18: 59.
 152. Murugaiyan J, Eravci M, Weise C, Roesler U (2017) Mass spectrometry data from label-free quantitative proteomic analysis of harmless and pathogenic strains of infectious microalgae, *Prototheca spp.* *Data in Brief* 12: 320–326.
 153. Murugaiyan J, Weise C, von Bergen M, Roesler U (2013) Twodimensional proteome reference map of *Prototheca zopfii* revealed reduced metabolism and enhanced signal transduction as adaptation to an infectious life style. *Proteomics* 13: 2664–2669.
 154. Muscatine L, McCloskey L, Marian R (1981) Estimating the daily contribution of carbon from zooxanthellae to coral animal respiration. *Limnology and Oceanography* 26: 601–611.
 155. Nakamura H, Kishi Y, Shimomura O, Morse D, Hastings J (1989) Structure of dinoflagellate luciferin and its enzymic and nonenzymic air-oxidation products. *Journal of the American Chemical Society* 111: 7607–7611.
 156. Nakamura M, Mie M, Funabashi H, Yamamoto K, Ando J, *et al.* (2006) Cell-surface-localized ATP detection with immobilized firefly luciferase. *Analytical Biochemistry* 352: 61–67.

157. Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford university press.
158. Ng S, Tan I, Kao H, Chang K, Chang S (2017) Recent developments on genetic engineering of microalgae for biofuels and bio-based chemicals. *Biotechnology Journal* 12: 1600644.
159. Ngo P, Mansoorabadi S (2017) Computational investigation of the mechanism of dinoflagellate bioluminescence: Chemically initiated electron exchange luminescence or twisted intramolecular charge transfer? *ChemPhotoChem* 1:9.
160. Nguyen H, Baudet M, Cuin_e S, Adriano J, Barthe D, Billon E, *et al.* (2011) Proteomic profiling of oil bodies isolated from the unicellular green microalga *Chlamydomonas reinhardtii*: with focus on proteins involved in lipid metabolism. *Proteomics* 11:4266–4273.
161. Nichols S, Roberts B, Richter D, Fairclough S, King N (2012) Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/beta-catenin complex. *Proceedings of the National Academy of Sciences of the United States of America* 109: 13046–13051.
162. Nicolas M, Nicolas G, Johnson C, Bassot J, Hastings J (1987) Characterization of the bioluminescent organelles in *Gonyaulax polyedra* (dinoflagellates) after fast-freeze fixation and antiluciferase immunogold staining. *Journal of Cell Biology* 105: 723-735.
163. Nymark M, Sharma K, Sparstad T, Bones M, Winge P (2016) A CRISPR/Cas9 system adapted for gene editing in marine algae. *Scientific Reports* 6: 24951.
164. Okamoto O, Asano, C, Aidar, E, Colepicolo P (1996) Of cadmium on growth and superoxide dismutase activity of this species of dinoflagellate. The marine microalga *Tetraselmis gracilis*. *Journal of Phycology* 32: 74-79.
165. Okamoto O, Hastings J (2003) Genome-wide analysis of redox-regulated genes in a dinoflagellate. *Gene* 4(321): 73-81.
166. Okamoto O, Liu L, Robertson D, Hastings J (2001) Members of a dinoflagellate luciferase gene family differ in synonymous substitution rates. *Biochemistry* 40: 15862-15868.
167. Okamoto, O.K.; Colepicolo, P (1998) Response of superoxide dismutase to pollutant metal stress in the marine dinoflagellate *Gonyaulax polyedra*. *Comp. Biochem. Physiol. C. Pharmacol. Toxicol. Endocrinol.* 119: 67-73.
168. Olaya-Abril A, Hidalgo-Carrillo J, Luque-Almagro V, Fuentes-Almagro C, Urbano F, Moreno-Vivián C, Richardson D, Roldán M (2018) Exploring the denitrification proteome of *Paracoccus denitrificans* PD1222. *Frontiers in Microbiology* 9: 1137.
169. Omata T. Murata N (1985) Electron-transport reactions in cytoplasmic and thylakoid membranes prepared from the cyanobacteria (blue-green algae) *Anacystis nidulans* and *Synechocystis* PCC 6714. *Biochimica et Biophysica Acta* 810: 354-361.
170. Orr R, Murray S, Stuken A, Rhodes L, Jakobsen K (2012) When naked became armored: an eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. *PLoS One* 7: e50004.
171. Palenik B, Grimwood J, Aerts A, Rouze P, Salamov A, *et al.* (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proceedings of the National Academy of Sciences of the United States of America* 104: 7705–7710.
172. Peschek G (1982) The cytochrome oxidase-hydrogenase relationship in cyanobacteria. *Naturwissenschaften* 69: 599-600.
173. Pierobon P, Cheng X, Graham P, Nguyen B, Karakolis E, Sinton D (2018) Emerging microalgae technology: A review. *Sustainable Energy Fuels* 2: 13–38.
174. Prevet A, Lindström J, Xu J, Karlson B, Selander E (2019) Grazer-induced bioluminescence gives dinoflagellates a competitive edge. *Current Biology* 29: R564-R565.
175. Priyadarshani P, Rath B (2012) Commercial and industrial applications of micro algae – A review. *Journal of Algal Biomass Utilization* 3(4): 89–100.
176. Read B, Kegel J, Klute M, Kuo A, Lefebvre S, Maumus F, *et al.* (2013) Pan genome of the phytoplankton *Emiliana underpins* its global distribution. *Nature* 499: 209–213.
177. Ren L, Chang E, Makky K, Haas A, Kaboord B, Walid Qoronfleh M (2003) Glutathione S-transferase pull-down assays using dehydrated immobilized glutathione resin. *Analytical Biochemistry* 322(2): 164-169.
178. Reynolds C, Islam S, Sternberg M (2018) EzMol: a web server wizard for the rapid visualization and image production of protein and nucleic acid structures. *J Mol Biol* 430: 2244-2248.
179. Rizzo P (1991) The enigma of the dinoflagellate chromosome. *Journal of Protozoology* 38: 246–252.

180. Rizzo P (2003) Those amazing dinoflagellate chromosomes. *Cell Research* 13: 215–217.
181. Rohr J, Sarkar N, Balenger S, Jeong BR, Cerutti H (2004) Tandem inverted repeat system for selection of effective transgenic RNAi strains in *Chlamydomonas*. *The Plant Journal* 40(4): 611–621.
182. Roustan V, Bakhtiari S, Roustan P, Weckwerth W (2017) Quantitative *in vivo* phosphoproteomics reveals reversible signaling processes during nitrogen starvation and recovery in the biofuel model organism *Chlamydomonas reinhardtii*. *Biotechnology for Biofuels* 10: 280.
183. Roy S, Morse D (2013) Transcription and maturation of mRNA in dinoflagellates. *Microorganisms* 1: 71–99.
184. Sasaki S, Drummen G, Konishi G (2016) Recent advances in twisted intramolecular charge transfer (TICT) fluorescence and related phenomena in materials chemistry. *Journal of Material Chemistry C* 4: 2731–2743.
185. Scaife A, Smith G (2016) Towards developing algal synthetic biology. *Biochemical Society Transactions* 44: 716–722.
186. Schenk P, Thomas-Hall S, Stephens E, Marx U, Mussgnug J, Posten C, *et al.* (2008) Second generation biofuels: high-efficiency microalgae for biodiesel production. *BioEnergy Research* 1: 20–43.
187. Schmitter R, Njus D, Sulzman F, Gooch V, Hastings J (1976) Dinoflagellate bioluminescence: A comparative study of *in vitro* components. *Journal of Cell Physiology* 87: 123–134.
188. Schmollinger S, Meuhlhaus T, Boyle NR, Blaby IK, Casero D, Mettler T, *et al.* (2014) Nitrogen-sparing mechanisms in *Chlamydomonas* affect the transcriptome, the proteome, and photosynthetic metabolism. *The Plant Cell* 26 (4): 1410–1435.
189. Schultz L, Liu L, Cegielski M, Hastings J (2005) Crystal structure of a pH-regulated luciferase catalyzing the bioluminescent oxidation of an open tetrapyrrole. *Proceeding of the National Academy of Sciences of the United States of America* 102:1378–83.
190. Scott A, Davey P, Dennis S, Horst I, Howe J, Lea-Smith J, *et al.* (2010) Biodiesel from algae: Challenges and prospects. *Current Opinion in Biotechnology* 21: 277–286.
191. Seo K, Fritz L (2000) Cell ultrastructural changes correlate with circadian rhythms in *Pyrocystis lunula* (Pyrrophyta). *Journal of Phycology* 36: 351–358.
192. Seo K, Fritz L (2001) Evidence for sexual reproduction in the marine dinoflagellates, *Pyrocystis noctiluca* and *Pyrocystis lunula* (Dinophyta). *Journal of Phycology* 37: 530–535.
193. Seo K, Fritz L (2006) Karyology of a marine non-motile dinoflagellate, *Pyrocystis lunula*. *Hydrobiologia* 563: 289–296.
194. Shao N, Bock R (2008) A codon-optimized luciferase from *Gaussia princeps* facilitates the *in vivo* monitoring of gene expression in the model alga *Chlamydomonas reinhardtii*. *Current Genetics* 53: 381–388.
195. Sharifiana S, Homaeib A, Hemmatic R, Khajehd K (2017) Light emission miracle in the sea and preeminent applications of bioluminescence in recent new biotechnology. *Journal of Photochemistry and Photobiology* 172: 115–128.
196. Shimomura O (1995) The roles of the two highly unstable components F and P involved in the bioluminescence of euphausiid shrimps. *Journal of Bioluminescence and Chemiluminescence* 10: 91–101.
197. Shimomura O (2006) *Bioluminescence: Chemical principles and methods*. World Scientific Publishing Co. Pte. Ltd., Singapore.
198. Shin S, Lim J, Koh H, Kim E, Kang N, Jeon S, *et al.* (2016) CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Scientific Reports* 6: 27810.
199. Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R, Takeuchi T, Hisata K, Tanaka M, Fujiwara M, *et al.* (2013) Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology* 23: 1399–1408.
200. Shukla P, Karthik K (2015) *Computational approaches in Chlamydomonas reinhardtii for effectual bio-hydrogen production*. Springer, New York.
201. Siegler H, Valerius O, Ischebeck T, Popko J, Tourasse NJ, Vallon O, *et al.* (2017) Analysis of the lipid body proteome of the oleaginous alga *Lobosphaera incisa*. *BMC Plant Biology* 17: 98.
202. Sigee D (1986) The dinoflagellate chromosome. *Advance of Botanical Research* 12: 205–264.

203. Singleton I (1994) Microbial metabolism of xenobiotics: Fundamental and applied research. *Journal of Chemical Technology and Biotechnology* 59(1): 9-23.
204. Sizova I, Greiner A, Awasthi M, Kateriya S, Hegemann P (2013) Nuclear gene targeting in *Chlamydomonas* using engineered zinc finger nucleases. *The Plant Journal* 73: 873–882.
205. Smith S, Morgan D, Musset B, Cherny V, Place A, Hastings J, Decoursey T (2011) Voltagegated proton channel in a dinoflagellate. *Proceedings of the National Academy of Sciences of the United States of America* 108: 18162–18167.
206. Soyer-Gobillard M (1996) The genome of the primitive eukaryote dinoflagellates; organization and functioning. *Zoological Study* 35: 78–84.
207. Spector D (1984) *Dinoflagellates*. Academic Press, Orlando, pp 545.
208. Stauber J, Binet M, Bao V, Boge J, Zhang A, Leung K, Adams M (2008) Comparison of the QwikLite™ algal bioluminescence test with marine algal growth rate inhibition bioassays. *Environmental Toxicology* 2008: 617-625.
209. Steidinger K, Tangen K (1997) *Dinoflagellates*. Tomas C (Ed.), *Identifying Marine Phytoplankton*. Academic Press, New York, pp. 519–520.
210. Surzycki R, Cournac L, Peltier G, Rochaix D (2007) Potential for hydrogen production with inducible chloroplast gene expression in *Chlamydomonas*. *Proceedings of the National Academy of Sciences of the United States of America* 104: 17548–17553.
211. Suzuki C, Nakajima Y, Akimoto H, Wu C, Ohmiya Y (2005) A new additional reporter enzyme, dinoflagellate luciferase, for monitoring of gene expression in mammalian cells. *Gene* 344: 61-66.
212. Swift E, Biggley W, Seliger H (1973) Species of oceanic dinoflagellates in genera *Dissodinium* and *Pyrocystis*: Interclonal and interspecific comparisons of color and photon yield of bioluminescence. *Journal of Phycology* 9: 420–426.
213. Swift E, Remsen C (1970). The cell wall of *Pyrocystis* spp. (*Dinococcales*). *Journal of Phycology* 6: 79–86.
214. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10: 512-526.
215. Tanikawa N, Akimoto H, Ogoh K, Chun W, Ohmiya Y (2004) Expressed sequence tag analysis of the dinoflagellate *Lingulodinium polyedrum* during dark phase. *Photochemistry and Photobiology* 80: 31-35.
216. Tesson B, Latz M (2015) Mechanosensitivity of a rapid bioluminescence reporter system assessed by atomic force microscopy. *Biophysical Journal* 108(6): 1341-1351.
217. Tinta T, Christiansen LS, Konrad A, Liberles DA, Turk V, Munch-Petersen B, Piškur J, Clausen A (2012) Deoxyribonucleoside kinases in two aquatic bacteria with high specificity for thymidine and deoxyadenosine. *FEMS Microbiology Letters* 331(2): 120-127.
218. Tokarev T, Vasilenko V, Mashukova O (2017) Some theoretical and applied aspects of the oceanic bioluminescence registration from space. *Advances in Bioscience and Biotechnology* 8: 283-294.
219. Topalov G, Kishi Y (2001) Chlorophyll catabolism leading to the skeleton of dinoflagellate and krill luciferins: Hypothesis and model studies. *Angewandte Chemie* 40: 3892–3894.
220. Toulza E, Shin M, Blanc G, Audic S, Laabir M, Collos Y, Claverie J, Grzebyk D (2010) Gene expression in proliferating cells of the dinoflagellate *Alexandrium catenella* (*Dinophyceae*). *Applied and Environmental Microbiology* 76: 4521–4529.
221. Uribe P, Fuentes D, Valdés J, Shmaryahu A, Zúñiga A, Holmes D, Valenzuela P (2008) Preparation and analysis of an expressed sequence tag library from the toxic dinoflagellate *Alexandrium catenella*. *Marine Biotechnology* 10: 692-700.
222. Vacher M, Galván I, Ding BW, Schramm S, Berraud-Pache R, Naumov P, Ferré N, Liu YJ, Navizet I, Roca-Sanjuán D, Baader W, Lindh R (2018) Chemo- and bioluminescence of cyclic peroxides. *Chemical Reviews* 118: 6927–6974
223. Valiadi M, Iglesias-Rodriguez D (2013) Understanding bioluminescence in dinoflagellates: how far have we come? *Microorganisms* 1: 3-25.
224. Valiadi M, Iglesias-Rodriguez D (2014) Diversity of the luciferin binding protein gene in bioluminescent dinoflagellates – Insights from a new gene in *Noctiluca scintillans* and sequences from Gonyaulacoid genera. *Journal of Eukaryotic Microbiology* 61(2): 134-145.
225. Valiadi M, Iglesias-Rodriguez D, Amorim A (2012) Distribution and genetic diversity of the luciferase gene within marine dinoflagellates. *Journal of Phycology* 48: 826-836.
226. Velazquez-Lucio J, Rodriguez-Jasso R, Colla L, Saenz-Galindo A, Cervantes-Cisneros D, Aguilar C, et al. (2018) Microalgal biomass pretreatment for bioethanol production: A review. *Biofuel Research Journal* 5(1): 780–791.

227. Veldhuis M, Cucci T, Sieracki M (1997) Cellular DNA content of marine phytoplankton using two new fluorochromes: Taxonomic and ecological implications. *Journal of Phycology* 33: 527-541.
228. Villarruel-Lopez A, Ascencio F, Nuno K (2017) Microalgae, a potential natural functional food source – A review. *Polish Journal of Food and Nutrition Sciences* 67 (4): 251–263.
229. Vizcaino JA, Csordas A, Del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, et al. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Research* 44: 11033.
230. Walsh C, Garneau-Tsodikova S, Gatto G (2005) Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* 44: 7342-7372.
231. Wang M, Liu Y (2017) Theoretical study of dinoflagellate bioluminescence. *Photochem Photobiol* 93: 511-518.
232. Wang Q, Lu Y, Xin Y, Wei L, Huang S, Xu J (2016) Genome editing of model oleaginous microalgae *Nannochloropsis* spp. by CRISPR/Cas9. *The Plant Journal* 88(6): 1071-1081.
233. Weretilnyk E, Alexander K, Drebenstedt M, Snider J, Summers P, Moffatt B (2001) Enhanced photosynthesis and redox energy production contribute to salinity tolerance in *Dunaliella* as revealed by homology-based proteomics. *Plant Physiology* doi:10.1104/pp.125.2.856.
234. Weyman P, Beerli K, Lefebvre S, Rivera J, McCarthy J, Heuberger A, et al. (2015) Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnology Journal* 13: 460–470.
235. Widder E (2001) Marine bioluminescence. *Biosci. Explain.* 1: 1–9.
236. Widder E (2010) Bioluminescence in the ocean: Origins of biological, chemical, and ecological diversity. *Science* 328: 704-708.
237. Widder E, Case J (1981) Bioluminescence excitation in a dinoflagellate. Nealson K (Ed.), *Bioluminescence Current Perspectives*. Burgess, Minneapolis, MN, pp. 125-132.
238. Widder E, Case J (1982) Distribution of sub-cellular bioluminescent sources in a dinoflagellate, *Pyrocystis fusiformis*. *Biology Bulletin* 162: 423-448.
239. Wilson T, Hastings J (1998) Bioluminescence. *Ann Rev Cell Dev Bi* 14: 197-230.
240. Wilson T, Hastings J (2013) *Bioluminescence: Living lights, lights for living*. Harvard University Press, Cambridge, MA, USA.
241. Wingett SW, Andrews S (2018) FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* 7.
242. Wisecaver J, Hackett J (2011) Dinoflagellate genome evolution. *Annual Review of Microbiology* 65: 369–387.
243. Worden A, Allen A (2010) The voyage of the microbial eukaryote. *Current Opinion in Microbiology* 13: 652–660.
244. Worden A, Lee J, Mock T, Rouze P, Simmons M, et al. (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 324: 268–272.
245. Wu C, Akimoto H, Ohmiya Y (2003) Tracer studies on dinoflagellate luciferin with [15N]-glycine and [15N]-L-glutamic acid in the dinoflagellate *Pyrocystis lunula*. *Tetrahedron Letters* 44: 1263–1266.
246. Yamaguchi A, Horiguchi T (2008) Culture of the heterotrophic dinoflagellate *Protoperdinium crassipes* (Dinophyceae) with noncellular food items. *Journal of Phycology* 44, 1090–1092.
247. Yamauchi Y, Hasegawa A, Taninaka A, Mizutani M, Sugimoto Y (2011) NADPH-dependent reductases involved in the detoxification of reactive carbonyls in plants. *Journal of Biological Chemistry* 286(9): 6999-7009.
248. Yang Z, Ma Y, Zheng J, Yang W, Liu J, Li H (2014) Proteomics to reveal metabolic network shifts towards lipid accumulation following nitrogen deprivation in the diatom *Phaeodactylum tricornutum*. *Journal of Applied Phycology* 26: 73–82.
249. Zhang H, Hou Y, Miranda L, Campbell D, Sturm N, Gaasterland T, Lin S (2007) Spliced leader RNA trans-splicing in dinoflagellates. *Proceedings of the National Academy of Sciences of the United States of America* 104, 4618–4623.
250. Zhang Z, Cavalier-Smith T, Green B (2002) Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins. *Molecular Biology and Evolution* 19: 489–500.
251. Zhang Z, Green B, Cavalier-Smith T (1999) Single gene circles in dinoflagellate chloroplast genomes. *Nature* 400: 155–159.

7. MATERIAL SUPLEMENTARIO

Supp. 1. Análisis de similitud de todos los conjuntos de datos por BLASTn. Resumen de las 3,645 búsquedas de BLASTn que comparan los 25 ensamblajes y dos conjuntos de datos públicos adicionales.

Supp. 2. Ensamble del transcriptoma generado con el software Trinity (Trinity_Q20 L100_k25_NC20), que mejor se alineó con los perfiles de identificación obtenidos por análisis MS.

Supp. 3. Resultado final de la anotación de proteoma después de todas las rondas de identificación a través de BLAST2GO.

Supp. 4. Análisis KEGG con todos los códigos y vías enzimáticos identificados en el proteoma de *P. lunula*.

Supp. 5. Nuevas proteínas en *P. lunula*. Secuencias que no tienen ninguna homología con ninguna proteína identificada hasta ahora en las diferentes bases de datos.

Supp. 6. Secuencia LBP obtenida por RT-PCR a partir de ARN de *P. lunula*.

Supp. 7. Secuencia LCF / LBP obtenida por RT-PCR a partir de ARN de *P. lunula*.

ANEXOS

INFORME DE INDICADORES DE CALIDAD DE PUBLICACIONES

Factor de Impacto y Cuartil ocupado por la revista según el Journal Citation Report, SCI y/o SSCI

1. Fajardo, C., De Donato, M., Carrasco, R., Martínez-Rodríguez, G., Mancera, J., Fernández-Acero, F. 2019. Advances and challenges in genetic engineering of microalgae. *Reviews in Aquaculture*, DOI: 10.1111/raq.12322

Factor de impacto (2018): 7,19; Cuartil: Q1, Ranking: 1/52 (Categoría: Fisheries)

2. Fajardo, C., Amil-Ruiz, F., Fuentes-Almagro, C., De Donato, M., Martínez-Rodríguez, G., Escobar-Niño, A., Carrasco, R., Mancera, J., Fernandez-Acero, F. 2019. An “omic” approach to *Pyrocystis lunula*: New insights related with this bioluminescent dinoflagellate., *Journal of Proteomics*, DOI: 10.1016/j.jprot.2019.103502

Factor de impacto (2019): 3,537; Cuartil: Q1

3. Fajardo, C., De Donato, M., Rodulfo, H., Martínez-Rodríguez, G., Costas, B., Mancera, J., Fernandez-Acero, F. 2019. New perspectives related to the bioluminescent system in dinoflagellates: *Pyrocystis lunula*, a case of study. *International Journal of Molecular Sciences*, DOI:10.3390/ijms21051784

Factor de impacto (2019): 4,183; Cuartil: Q1